
Mixed *Ab Initio* QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide

DEAN M. PHILIPP, RICHARD A. FRIESNER

Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York, 10027

Received 23 October 1998; accepted 10 May 1999

ABSTRACT: Methodology is discussed for mixed *ab initio* quantum mechanics/molecular mechanics modeling of systems where the quantum mechanics (QM) and molecular mechanics (MM) regions are within the same molecule. The *ab initio* QM calculations are at the restricted Hartree–Fock level using the pseudospectral method of the Jaguar program while the MM part is treated with the OPLS force fields implemented in the IMPACT program. The interface between the QM and MM regions, in particular, is elaborated upon, as it is dealt with by “breaking” bonds at the boundaries and using Boys-localized orbitals found from model molecules in place of the bonds. These orbitals are kept frozen during QM calculations. Results from tests of the method to find relative conformational energies and geometries of alanine dipeptides and alanine tetrapeptides are presented along with comparisons to pure QM and pure MM calculations. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 1468–1494, 1999

Keywords: *ab initio*; quantum mechanics; molecular mechanics; QM/MM; force field; peptides; proteins; enzyme active-site; computer-simulation

Introduction

During the past decades, there have been numerous attempts to develop computational methods for combining quantum-mechanical (QM)

and molecular-mechanical (MM) representations of molecules.¹ The goal of these efforts has been to be able to treat a small part of a large system via QM (e.g., the core region of a transition-state structure, or an enzyme-active site) and the remainder of the system via MM, thus obtaining high accuracy where it is required while avoiding the computational demands of QM calculations for a large number of atoms.

When the QM and MM subsystems are not connected by a chemical bond (e.g., if one were to

Correspondence to: R. A. Friesner; e-mail: rich@chem.columbia.edu

Contract/grant sponsor: NIH; contract/grant numbers: GM-40526, RR0682

treat a solute via QM and a solvent via MM,²⁻⁷ it is relatively straightforward to develop a formalism for QM/MM calculations; the electrostatic energy is easily computed via a Coulomb interaction between the MM point charges and the QM wave function, whereas the remaining nonbonded interaction between QM and MM atoms is represented by a van der Waals interaction (the parameters of which may have to be optimized). However, when the QM and MM regions are joined by one or more chemical bonds (as is the case, e.g., in an enzyme-active site), substantial formal and practical difficulties are immediately manifested. This is particularly the case if one intends to allow the MM charges to polarize the QM region, a feature that is one of the principal reasons for carrying out QM/MM calculations in the first place.

A number of approaches have been developed to solve this problem. The great majority of methods proposed to date begin by capping the QM and MM fragments with hydrogens. In what follows, we shall refer to the hydrogen-capped QM fragment as *A*, the MM fragment as *B*, and the entire coupled system as *S*. At this point, the simplest methodology, for example, as that used by Thiel and coworkers and by Morakuma and coworkers, is to carry out a QM calculation for the QM fragment *A* and MM calculations for both the QM fragment and for the entire system.^{8,9} Defining the MM energy of the whole system as S_{MM} , the MM energy of the QM fragment as A_{MM} , and the QM energy of the QM fragment as A_{QM} , the total QM/MM energy of the system is then written as:

$$E_{\text{QM/MM}} = S_{\text{MM}} + (A_{\text{QM}} - A_{\text{MM}}) \quad (1)$$

This approach has the great advantage of simplicity (indeed, it is completely trivial to implement computationally), and plausibly captures many of the desired characteristics of a useful QM/MM methodology. However, it has several deficiencies that make it potentially unsuitable for the full range of possible applications of QM/MM technology. First, there may not be a suitable MM representation of the QM fragment available (e.g., a transition state), and developing one could be very time-consuming. Second, one cannot treat electrostatic polarization of the QM region by the MM charge distribution with this approach. Third, termination with a hydrogen is probably safe from an electronic structure point of view only if the atom to which it is attached is sufficiently far from the reactive core of the problem; in a three-dimen-

sional system, this may lead to the requirement of a quite large QM region. For the applications we wish to pursue, modeling of transition metal active sites of metalloprotein enzymes, all of these objections are serious issues.

A second class of method attempts to include electrostatic polarization in the context of a hydrogen-capping method, while at the same time avoiding the necessity of having available a MM representation of the QM region (other than van der Waals interactions).^{2,8,10-17} These methods, which generally involve the use of "link atoms," face formidable problems in properly describing the QM/MM boundary in terms of both electrostatics and valence terms.¹¹ We do not discuss the details of these methods here, as they are not particularly relevant to our own approach. Although significant progress has been made in improving link atom methods, it is not yet the case that demonstrably reliable, high-precision results are routinely available.

In the present study, we pursue a different direction, based on the use of localized frozen quantum-chemical orbitals, as a means of creating an interface between the QM and MM regions of the molecule.^{1,18,19} Although there has been some previous work along these lines, the results presented therein were very preliminary for QM at the *ab initio* level.²⁰ The *ab initio* QM/MM method presented here is far more detailed and allows for the movement of atoms at the QM/MM interface. The fact that these interfacial atoms are allowed to move means that the frozen orbitals must be allowed to change with changes in geometry, and thus there are special challenges in this implementation, including the calculation of energy gradients with respect to atomic coordinates. As in the case of link atoms, there are formidable technical difficulties in developing a robust, accurate frozen orbital methodology. What we report, while far from being optimized, does represent a reasonable first generation solution to these difficulties.

One of the most difficult tests for a QM/MM scheme is the reproduction of conformational energies for a polar system. Even in a nonpolar system, there can be significant problems if the molecular mechanics force field contains partial atomic charges, as Brooks and coworkers have demonstrated for the torsional profile of butane.¹¹ We choose here to study alanyl peptides, which undoubtedly represent a significant challenge for any QM/MM methodology. Our QM/MM interface parameters were developed for the alanine dipep-

tide, then used to compute conformational energetics of the alanine tetrapeptide, which are compared with fully quantum-mechanical calculations. To our knowledge, no tests of this type (other than the work on butane alluded to earlier and in other sources,^{21,22} a qualitatively simpler system) have been reported in the literature for alternative QM/MM methods. The relative energetics of the conformers are highly dependent on having an accurate description of the energetics at the QM/MM interface, and hence constitute a severe test for the methodology. In contrast, other measures, such as interaction of the QM part of the system with a small molecule, are relatively trivial, as the errors in the energetics at the interface (if it is sufficiently far removed from the targeted interactions) will cancel. It is our view that methods that cannot reproduce conformational energetics will either require very large QM regions or be of limited utility.

The price that has to be paid for achieving reasonable accuracy in the conformational energetics is the investment of significant effort in QM/MM interface parameterization. The scheme we present here is probably not the best one, but it appears to provide an accuracy comparable to that of the molecular mechanics itself, a reasonable initial goal for a QM/MM methodology. Our intention is to focus initially on QM/MM modeling of peptides and proteins, and to develop interfacial parameters specific to this application. In general, we expect that QM/MM parameter development will be just as demanding and time consuming as the development of MM force fields—if anything, the interface problem is more difficult than that of pure MM parameterization.

This article is organized as follows. The next section presents an overview of our frozen orbital method, and outlines the basic strategy for development of the model. The third section describes in detail our QM/MM methodology using frozen orbitals. The fourth section presents computational details and results for relative conformational energetics of the alanine dipeptide and tetrapeptide, providing a comparison with standard quantum-mechanical and molecular-mechanics results. The Conclusion section discusses future directions.

Overview and Basic Strategy

The method presented here will be used for such applications as single-point energy calculations and geometry optimizations for large sys-

tems where a small part is of special interest, such as the active site of an enzyme. Because the present research is concerned with the case in which a single molecule is divided into QM and MM regions, how the molecule is partitioned is the first concern. Once the molecule has been partitioned, the idea of the method used is to place localized orbitals at the interfaces between the QM and MM parts of the molecule; these orbitals are to remain frozen during calculations for the QM part of the molecule with a given geometry. These frozen orbitals are to be obtained from QM calculations for model molecules, with the strategy for practical modeling applications to involve building a database of such orbitals. The MM part of the molecule will be treated with MM force fields and the QM part with the desired level of quantum mechanics. Interactions between the QM and MM regions will need to be defined, with the ability of the MM region to polarize the QM region. The total system energy, as well as energy gradients with respect to atomic coordinate changes, will thus be comprised of contributions from the QM part, the MM part, and interactions between the two. The gradients will be necessary when performing geometry optimizations, as energies and gradients will need to be calculated at each step to determine new geometries until a minimum geometry is converged upon.

Methodology

PARTITIONING OF QM/MM MOLECULE

First, the molecule is “cut” at a particular bond into two parts. Care must be taken in choosing the bond at which the molecule is divided as it should be far enough away from the part of the molecule of special interest (e.g., the active site of an enzyme), and it should not be at a bond with a significant degree of multiple bonding character. However, the actual guidelines to follow in choosing the dividing bond represent an empirical matter and need to be addressed for the specific type of molecule in question. For work with peptides and proteins, the two types of bonds along the backbone considered as possible sites for dividing the molecule are the bond between the α -carbon and the amino nitrogen and the bond between the α -carbon and the carboxyl carbon. Another good candidate for cutting is along the side chain between the α -carbon and β -carbon.

Once the dividing bond is chosen, all atoms on the side, including the part of the molecule of special interest, constitute the QM region, whereas all atoms on the other side are part of the MM region. The atoms in the MM region will be replaced by point charges, and interactions between these atoms will be represented by MM potential energy terms. The QM region will be treated with *ab initio* quantum chemistry with the interactions being the normal Hamiltonian terms. There will also be interactions between the QM and the MM regions, but more will be said about all these terms later. The bond at the site of the division will be replaced by a frozen orbital (as described in the next subsection), and will therefore be referred to as a frozen bond. Multiple frozen bonds can be chosen; the frozen bonds then separate the QM region from various unconnected parts of the molecule that together compose the MM region.

FORMATION OF FROZEN ORBITALS

As mentioned earlier, frozen orbitals will be used at the sites of the frozen bonds. Each of these orbitals is a spatial Hartree–Fock molecular orbital found from the converged wave function of a (relatively) small molecule and will be referred to as the model molecule. This model molecule will need to be similar to the actual QM/MM molecule in that it must contain the molecular fragment surrounding the frozen bond. This means that the model molecule should at least contain the atoms adjacent to the frozen bond and all the atoms directly connected to these two atoms, but how far beyond this is an empirical question. More atoms will need to be included if there are important functional groups adjacent to the frozen bond; however, increasing the size of the model molecule will naturally lead to longer computational times. The idea then is to use the smallest model molecule possible that yields a frozen orbital that resembles well enough what the orbital would look like in the actual QM/MM molecule. Note also that using a larger model molecule will make a particular frozen orbital applicable to fewer QM/MM molecules, meaning that a larger database of frozen orbitals will need to be formed. Once all the necessary atoms for the model molecule are specified, any unfulfilled valences beyond these atoms are terminated with hydrogens.

Once the converged wave function for the model molecule is found, the occupied molecular orbitals are Boys-localized.²³ The Boys-localized orbital with significant electron density on the two atoms

adjacent to the frozen bond, indicating that it is the “ σ -orbital” for the frozen bond, is the frozen orbital. Further manipulations must be made to this frozen orbital before it can be used in the QM/MM molecule, but this is the starting point for all frozen bonds to which it is applied.

USE OF FROZEN ORBITALS IN QM/MM MOLECULE

The first manipulation of the frozen orbital when placing it in the QM/MM molecule is to remove all basis functions centered on atoms other than the two atoms adjacent to the frozen bond and all the atoms connected to these two atoms (see Fig. 1). This means zeroing out the coefficients for all these basis functions. The requirement that the coefficients for the removed basis functions be small is obvious, which should indeed be the case for the localized frozen orbitals.

The next operation on the frozen orbital is to rotate it so that it is lined up along the axis of the frozen bond in the QM/MM molecule. This is done by choosing one of the atoms connected to one of the two atoms adjacent to the frozen bond as a reference atom (this atom cannot be collinear with the two bond-adjacent atoms). Then the necessary rotation matrix is found to match the orientation of the two adjacent atoms and the reference atom in the model molecule with the respective atoms in the QM/MM molecule. The resulting orbital rotation matrix, R , is then applied to the basis functions of the frozen orbital:

$$\mathbf{c}' = R\mathbf{c} \quad (2)$$

where \mathbf{c} and \mathbf{c}' are the coefficient vectors for the frozen orbital before and after rotation.

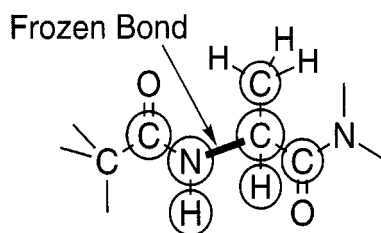


FIGURE 1. Removal of basis functions from a frozen orbital. The basis functions removed are those centered on atoms other than the two atoms adjacent to the frozen bond and all the atoms connected to these two atoms (circled atoms). This is accomplished by zeroing out the coefficients for these removed basis functions in the frozen orbital.

In addition to this general orbital rotation, azimuthal rotations of the basis functions centered on the atoms connected to the two frozen bond-adjacent atoms may be necessary if the torsional angles of these atoms in the reference frame formed by the two bond-adjacent atoms and the reference atom in the model molecule are different than the respective torsional angles in the reference frame of the QM/MM molecule. The relative torsional angles, ϕ_i , are thus calculated for all the atoms, i , connected to the two bond-adjacent atoms (it is not necessary to do this for the reference atom), and then the basis functions on each of these atoms are rotated by the respective rotation matrices, $\mathbf{R}_{\text{tor}}(\phi_i)$ (these azimuthal rotations are actually included within the total rotation matrix \mathbf{R}). A demonstration of this process to determine the rotations is shown in Figure 2, and a detailed discussion of how the actual rotation matrices are found is in the Appendix.

After the frozen orbitals are rotated so that they are properly aligned in the QM/MM molecule, they must be re-orthonormalized. This is necessary not only because of the removal of some of the basis functions, but also because the overlap between basis functions located on different atoms will differ as distances and angles between atoms

change. This re-orthonormalization can be represented by the following matrix transformation:

$$\mathbf{C}' = \mathbf{C}\mathbf{U}\boldsymbol{\sigma} \quad (3)$$

where \mathbf{C} and \mathbf{C}' are the coefficient matrices for the frozen orbitals before and after re-orthonormalization, respectively. \mathbf{U} is a unitary matrix, and $\boldsymbol{\sigma}$ is a diagonal matrix ($\mathbf{C}'\boldsymbol{\sigma}^{-1}\mathbf{U}^+$ is the singular value decomposition of \mathbf{C}). Combining eqs. (2) and (3) yields the following transformation:

$$\mathbf{C}' = \mathcal{R}\mathbf{C}\mathbf{U}\boldsymbol{\sigma} \quad (4)$$

where \mathcal{R} is now a tensor of rank three due to the fact that there is a different rotation matrix \mathbf{R}^i for each frozen orbital \mathbf{c}_i in \mathbf{C} .

Once the frozen orbitals have been transformed for use in the QM/MM molecule, they will need to remain frozen throughout the Hartree-Fock self-consistent field (SCF) calculations. To do this, the normal Roothaan equation must be modified to the following equation:

$$\mathbf{F}''\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\epsilon}, \text{ or } \sum_{\nu} F''_{\mu\nu} C_{\nu i} = \sum_{\nu} \epsilon_i S_{\mu\nu} C_{\nu i} \quad (5)$$

$$F''_{\mu\nu} = F_{\mu\nu} - \sum_{\substack{i=\text{non} \\ j=\text{fr}}} \sum_{\alpha\beta} S_{\mu\alpha} C_{\alpha i} \epsilon_{ij} C_{\beta j} S_{\beta\nu} \quad (6)$$

Here, \mathbf{F} is the usual Fock matrix, \mathbf{F}'' is the modified Fock matrix, \mathbf{S} is the overlap matrix, and $\boldsymbol{\epsilon}$ is the matrix of orbital energies, which is now no longer diagonal, because $\epsilon_{ij}, j \neq i \neq 0$ when i and/or j is a frozen orbital. Note that, in eq. (6), the first summation is over $i = \text{nonfrozen orbitals (non)}$ and $j = \text{frozen orbitals (fr)}$. The modified Roothaan equations insure that the frozen orbitals remain the same, as they are eigenvectors of these equations and are prevented from mixing with any other orbitals. The complete derivation of these equations can be found in the Appendix.

Finally, electron density must not be permitted to flow from the nonfrozen orbitals to the MM region. The electron repulsion from the frozen orbitals and the fact that the nonfrozen orbitals must remain orthogonal to the frozen orbitals will only prevent electron density from transferring to the immediate vicinity of the frozen orbitals. Therefore, the nonfrozen orbitals must not be allowed to populate basis functions centered on atoms in the MM region (which amounts to the MM atoms adjacent to the frozen bonds and the MM atoms connected to these atoms; basis functions centered on any other MM atoms are not used for any orbitals). To accomplish this, the initial guess for

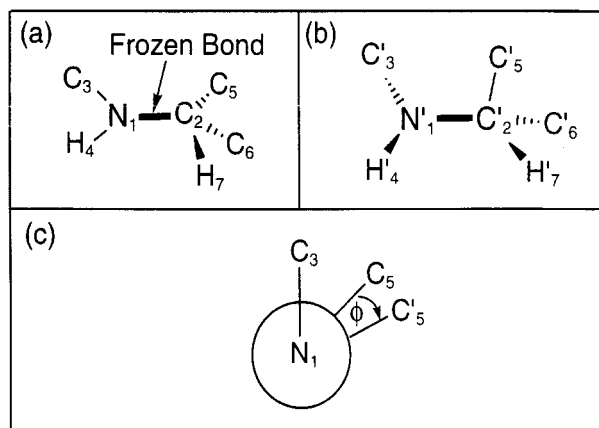


FIGURE 2. Demonstration of how rotations for frozen orbitals are determined. (a) Reference frame of the model molecule, and (b) reference frame of the QM/MM molecule (primed reference frame). Euler angles align $\text{C}_3-\text{N}_1-\text{C}_2$ with $\text{C}'_3-\text{N}'_1-\text{C}'_2$. Looking down the N_1-C_2 frozen bond axis, (c) shows the relative torsional angle ϕ for which C_5 must be rotated to be aligned with C'_5 . The relative torsional angles ϕ_i are found for atoms H_4 , C_5 , C_6 , and H_7 (it is not necessary to find ϕ for C_3 because it is the reference atom); and then the azimuthal rotations of the basis functions centered on these respective atoms are incorporated into the orbital rotation matrix \mathbf{R} .

the nonfrozen occupied orbitals, which is the semiempirical initial guess formulated for the molecular fragment of the QM/MM molecule that contains only the QM-region atoms, is made to be orthonormal to the frozen orbitals and to each other without the use of the forbidden basis functions. Currently, this orthonormalization procedure uses the modified Gram–Schmidt process, which involves altering all vectors following the given vector at each step so that these remaining vectors are made to be orthogonal to this given vector. When making the nonfrozen orbitals orthogonal to each frozen orbital, only the projection of that frozen orbital onto the space of the allowed basis functions (i.e., the vector comprised of just the coefficients for the allowed basis functions) is multiplied by a scalar and added to the nonfrozen orbitals to yield the new orthogonal orbitals. Then, as many of the virtual orbitals as can remain linearly independent are made to be orthonormal without the use of the forbidden orbitals. The remaining virtual orbitals are allowed to populate all of the basis functions as they are made to be orthonormal. Keeping these virtual orbitals frozen (thus they will be referred to as frozen virtuals) and preventing them from mixing with any other orbitals in the same manner as the frozen orbitals will bring about the desired result of preventing the nonfrozen orbitals from populating the forbidden basis functions. So, as for the frozen orbitals, the frozen virtuals are made to be eigenvectors of a modified Roothaan equation:

$$\mathbf{F}'\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\epsilon}, \text{ or } \sum_{\nu} F'_{\mu\nu} C_{\nu i} = \sum_{\nu} \epsilon_i S_{\mu\nu} C_{\nu i} \quad (7)$$

$$F'_{\mu\nu} = F_{\mu\nu} - \sum_{\substack{i=\text{non} \\ j=\text{fr}}} \sum_{\alpha\beta} S_{\mu\alpha} C_{\alpha i} \epsilon_{ij} C_{\beta j} S_{\beta\nu} - \sum_{\substack{i=\text{non} \\ j=\text{fr. virt.}}} \sum_{\alpha\beta} S_{\mu\alpha} C_{\alpha i} \epsilon_{ij} C_{\beta j} S_{\beta\nu} \quad (8)$$

These are similar to eqs. (5) and (6), except that eq. (8) has a second term where the first summation is over $i = \text{nonfrozen orbitals}$ and $j = \text{frozen virtuals (fr. virt.)}$. These are the equations that must be solved during the SCF calculations.

DETERMINATION OF SYSTEM ENERGY

The total system energy for the QM/MM molecule consists of contributions from the QM region, MM region, and interactions between the QM and MM regions. The terms for the QM region

involve the following:

$$E_{\text{QM}}^* = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{\text{core}*} + \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} [2J_{\mu\nu} - K_{\mu\nu}] + \sum_{AC} \frac{Z_A Z_C}{R_{AC}} + \sum_{AM} \frac{Z_A q_M}{r_{AM}} \quad (9)$$

where the * indicates that the QM/MM electrostatic interactions have now been included here, and q_M is the partial charge of MM atom M . The JAGUAR²⁴ program at the restricted HF level performs the actual SCF calculations to determine the ground-state wave function and energy in the presence of the MM point charges and with the use of the frozen orbitals at the frozen bonds.

The partial charges assigned to the MM atoms, as seen by the QM region (and the rest of the MM region) through electrostatic interactions, are the same charges that would be assigned by the MM force field used (OPLS in the case of this research), with the exception of the MM atom that is adjacent to the frozen bond. Initial attempts at proper assignment of this charge involved starting with the charge that the force field would assign to it and add whatever charge was necessary so that the QM/MM molecule was electrically neutral. This meant that, because the QM region, including the frozen orbital, would add up to a total charge of -1 , the charge on the MM atom adjacent to the frozen bond would be adjusted so that all MM partial charges would sum to $+1$.

This scheme, however, did not yield an electrostatic potential similar enough to the actual potential that would be experienced by the QM region if the MM region were treated quantum mechanically. Therefore, the current protocol is to include an additional MM point charge at the midpoint of the frozen bond and to make this charge plus the charge on the MM atom adjacent to the frozen bond add up to the charge necessary to make the QM/MM molecule neutral. This offers a degree of freedom (there is potentially another degree of freedom as the position of this “bond charge” could be varied instead of just being placed at the bond midpoint) to produce an electrostatic potential that yields more desirable results. A rigorous method for determining the bond charge has not been developed as of yet; rather, the bond charge is determined by trial and error so that energies of deprotonation for QM hydrogens far enough away from the QM/MM interface (at least three bonds away from the QM bond-adjacent atom) are reasonably reproduced through the QM/MM model-

ing as compared with full QM. For these energies of deprotonation, the geometries are simply kept frozen while the relevant proton is removed. As the geometries are kept frozen, the energies of deprotonation do not depend on any of the other necessary QM/MM parameters (described later), and thus the assignment of the bond charge can be done prior to fitting the other parameters. Test charges have also been used to verify the proper assignment of the bond charge. The resulting bond charge will also interact with the MM partial charges in the same manner as other MM charges, although some of the interactions with nearby atoms will be neglected or scaled, as discussed in what follows with the QM/MM correction energies.

The MM terms include the normal bonding terms for bond stretching, bond angle bending, and torsional twisting, and nonbonding terms for Coulomb and van der Waals interactions:

$$\begin{aligned}
 E_{\text{MM}} = & \sum_{i=\text{stretches}} k_i (r_i - r_0)^2 + \sum_{j=\text{bends}} k_j (\theta_j - \theta_0)^2 \\
 & + \sum_{k=\text{torsions}} \left[\frac{V_{k,1}}{2} (1 + \cos \phi_k) \right. \\
 & + \frac{V_{k,2}}{2} (1 - \cos 2\phi_k) + \left. \frac{V_{k,3}}{2} (1 + \cos 3\phi_k) \right] \\
 & + \sum_{MN} \frac{q_M q_N}{r_{MN}} \\
 & + \sum_{MN} 4\epsilon_{MN} \left[\left(\frac{\sigma_{MN}}{r_{MN}} \right)^{12} - \left(\frac{\sigma_{MN}}{r_{MN}} \right)^6 \right] \quad (10)
 \end{aligned}$$

where the torsions include both proper and improper torsions and the M and N labels refer to MM atoms. The first three terms are the bonding terms, whereas the last two terms are nonbonding. The nonbonding terms are normally scaled by 0.5 for 1–4 interactions (interactions for atoms that are three bonds away), and they are normally not included for 1–2 and 1–3 interactions (interactions that are one and two bonds away, respectively). The actual energy calculations are performed by the IMPACT²⁵ program, which employs OPLS^{26,27} force fields.

The interactions between the QM and MM regions include electrostatic terms between the MM point charges and the QM nuclei and electrons (these, as explained earlier, are actually included with the pure QM terms), van der Waals terms between MM atoms and QM atoms, and MM-like terms at the intersection of the QM and MM re-

gions. These last terms are correction terms that account for the fact that the electrostatic and van der Waals terms between the QM and MM atoms are insufficient to represent differences in energy when the atoms at the QM/MM boundary are moved. Also a factor in these correction terms is the energy discrepancy resulting from the frozen orbitals only being rotated and re-orthonormalized instead of having their coefficients accurately adjusted as the relevant atomic coordinates involved change.

The first of these correction terms at the interface are terms that attempt to reduce or eliminate the short-range electrostatic interactions in terms of the energy. This is more consistent with the MM practice neglecting 1–2 and 1–3 electrostatic interactions and scaling the 1–4 interactions by 0.5, and attempts to remove some of the spurious short-range interactions that would make the remaining bonding-like terms too difficult to cast into a simple form. Because it is undesirable to remove completely any of the MM charges from acting on the QM wave function in the Hamiltonian, and impractical to only remove and/or scale certain interactions in the Hamiltonian (this would also lead to great difficulty in calculating the gradients as they no longer could be computed analytically), the form of these electrostatic (e.s.) correction terms is:

$$E_{\text{e.s. corr}} = - \sum_M \sum_Q \sigma_{MQ} \frac{q_M q_Q^*}{r_{MQ}} \quad (11)$$

Here M and Q are MM and QM atoms, respectively, σ_{MQ} is a scaling factor that is 1.0 for a 1–2 or 1–3 interaction, 0.5 for a 1–4 interaction (except when M is the bond adjacent MM atom, in which case $\sigma_{MQ} = 0.0$), and 0.0 for all others, and q_Q^* is the MM-like partial charge assigned to atom Q that approximates the actual charge that the MM atom experiences. The q_Q^* values are thus assigned to the same charges that would be assigned by the MM force field (again, OPLS here), with the exception of the bond-adjacent QM atom. It is assigned to a charge such that the sum of the q_Q^* is +1 (the total charge on the QM region minus the two electrons in the frozen bond is +1).

Additional electrostatic correction terms involve the frozen orbitals and bond charges. Interactions between the frozen orbitals and the MM charges can be computed explicitly both in the energy and gradient expressions, so scaling or removing such terms is possible. The bond charges (b.c.) interact with the MM charges just like the other MM charges, whereas scaling or removal of interactions

between the bond charge and the neighboring part of the QM region is analogous to the scaling and removal of interactions between the MM charges and the QM region as just demonstrated. The total (and now complete) electrostatic correction energy can thus be represented as:

$$\begin{aligned}
 E_{\text{e.s. corr}} = & - \sum_M \sum_Q \sigma_{MQ} \frac{q_M q_Q^*}{r_{MQ}} \\
 & - 2 \sum_M \sum_{a=\text{fr}} \sigma_{Ma} \left\langle a \left| \frac{q_M}{r_{Ma}} \right| a \right\rangle \\
 & - \sum_M \sum_{B=\text{b.c.}} \sigma_{MB} \frac{q_M q_B}{r_{MB}} \\
 & - \sum_Q \sum_{B=\text{b.c.}} \sigma_{QB} \frac{q_Q^* q_B}{r_{QB}} \\
 & - 2 \sum_{B=\text{b.c.}} \sum_{a=\text{fr}} \sigma_{Ba} \left\langle a \left| \frac{q_B}{r_{Ba}} \right| a \right\rangle \quad (12)
 \end{aligned}$$

Exactly what σ_{Ma} , σ_{MB} , σ_{QB} , and σ_{Ba} are, however, is an empirical matter. After observing how the total energy calculated through modeling of a QM/MM molecule (alanine dipeptide and alanine tetrapeptide conformers have been used to this point) for various perturbed geometries compares with the total energy calculated through pure QM, the following rules for the scalings were ascertained:

- $\sigma_{Ma} = 1.0$ if M is the bond-adjacent MM atom or an MM atom connected directly to the bond-adjacent MM atom, $\sigma_{Ma} = 0.75$ if M is two bonds from the bond-adjacent MM atom, $\sigma_{Ma} = 0.25$ if M is three bonds away, and $\sigma_{Ma} = 0.0$ for all other M .
- σ_{MB} follows the same rules as σ_{Ma} .
- $\sigma_{QB} = 1.0$ if Q is the bond-adjacent QM atom or a QM atom connected directly to the bond-adjacent QM atom and $\sigma_{QB} = 0.0$ for all other Q .
- $\sigma_{Ba} = 1.0$ if B is on the same frozen bond as a and $\sigma_{Ba} = 0.0$ for all other cases.

Note that these rules assume that all frozen bonds are far enough removed from one another.

The remaining correction terms are MM-like bonding terms, and include stretch terms, bend terms, and torsion terms. The stretch terms are between a pair of atoms where at least one of the two atoms are one of the bond-adjacent atoms, whereas the bend terms are for all bond angles for which the two bond-adjacent atoms are two of the

three atoms involved in the bond angle. The torsion terms include all torsional angles in which the bond-adjacent atoms are two of the four atoms involved. However, if the QM bond-adjacent atom is one of the end atoms, then the torsional terms simply use the same parameters used by the employed force field.

The parameters for the stretch and bend correction terms are found by first perturbing the minimum geometry (or minimum geometries if multiple minima exist for a model molecule and it is deemed necessary to use more than one of these) for the model molecule. Then the resulting QM/MM energies are compared with the pure QM energies and the energy differences are fit to the relevant stretch and bend terms. The stretch and bend terms are of the form:

$$E_{\text{str. corr.}} = \sum_i k_i (r_i - r_{i0})^2 \quad (13)$$

$$E_{\text{bend corr.}} = \sum_j k_j (\theta_j - \theta_{j0})^2 \quad (14)$$

where k_i , r_{i0} , θ_j , and θ_{j0} are the fit parameters. The fitting procedure is simply a linear least-squares fit using singular value decomposition to first fit the energy differences (plus an arbitrary zero-point energy constant) to expressions of the form:

$$E'_{\text{str. corr.}} = \sum_i \alpha_i (r_i - r_0) + k'_i (r_i - r_0)^2 \quad (15)$$

$$E'_{\text{bend corr.}} = \sum_j \alpha_j (\theta_j - \theta_0) + k'_j (\theta_j - \theta_0)^2 \quad (16)$$

where α_i , k'_i , α_j , and k'_j are the fit parameters and r_0 and θ_0 are some reference bond length and bond angle. The desired parameters are then related to these parameters as:

$$k_i = k'_i, \quad r_{i0} = r_0 - \frac{\alpha_i}{2k'_i} \quad (17)$$

$$k_j = k'_j, \quad \theta_{j0} = \theta_0 - \frac{\alpha_j}{2k'_j} \quad (18)$$

The parameters for the torsion correction terms are determined by first varying the relevant torsional angles for which parameters are desired and then executing restricted optimizations where these same torsional angles are kept frozen. Then the resulting QM/MM energies (which include the stretch and bend corrections just found) are compared with the pure QM energies, and the differ-

ences are fit with the torsional expressions:

$$E_{\text{tor. corr.}} = \sum_k \left[\frac{V_{k,1}}{2} (1 + \cos \phi_k) + \frac{V_{k,2}}{2} (1 - \cos 2\phi_k) + \frac{V_{k,3}}{2} (1 + \cos 3\phi_k) \right] \quad (19)$$

Here, $V_{k,1}$, $V_{k,2}$, and $V_{k,3}$ are the fit parameters found by performing a linear least-squares fit using singular value decomposition. The least-squares fit can be weighted so that geometries near the minimum (or minima) are weighted more heavily in the fit.

Now that all the individual components have been discussed, eqs. (9), (10), (12), (13), (14), and (19) can be combined to give a master equation for the system energy:

$$\begin{aligned} E_{\text{tot}} = & \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{\text{core}*} + \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} [2J_{\mu\nu} - K_{\mu\nu}] \\ & + \sum_{AC} \frac{Z_A Z_C}{R_{AC}} + \sum_A \sum_{B=\text{b.c.}} \frac{Z_A q_B}{r_{AB}} \\ & + \sum_{AM} \frac{Z_A q_M}{r_{AM}} + \sum_{i=\text{stretches}} k_i (r_i - r_0)^2 \\ & + \sum_{j=\text{bends}} k_j (\theta_j - \theta_0)^2 \\ & + \sum_{k=\text{torsions}} \left[\frac{V_{k,1}}{2} (1 + \cos \phi_k) + \frac{V_{k,2}}{2} (1 - \cos 2\phi_k) + \frac{V_{k,3}}{2} (1 + \cos 3\phi_k) \right] \\ & + \sum_{Mn} \frac{q_M q_N}{r_{MN}} + \sum_M \sum_{B=\text{b.c.}} \frac{q_M q_B}{r_{MB}} \\ & + \sum_{MN} 4\epsilon_{MN} \left[\left(\frac{\sigma_{MN}}{r_{MN}} \right)^{12} - \left(\frac{\sigma_{MN}}{r_{MN}} \right)^6 \right] \\ & - \sum_M \sum_Q \sigma_{MQ} \frac{q_M q_Q^*}{r_{MQ}} - 2 \sum_M \sum_{a=\text{fr}} \sigma_{Ma} \langle a | \frac{q_M}{r_{Ma}} | a \rangle \\ & - \sum_M \sum_{B=\text{b.c.}} \sigma_{MB} \frac{q_M q_B}{r_{MB}} - \sum_Q \sum_{B=\text{b.c.}} \sigma_{QB} \frac{q_Q^* q_B}{r_{QB}} \\ & - 2 \sum_{B=\text{b.c.}} \sum_{a=\text{fr}} \sigma_{Ba} \langle a | \frac{q_B}{r_{Ba}} | a \rangle \\ & + \sum_i k_i (r_i - r_{i0})^2 + \sum_j k_j (\theta_j - \theta_{j0})^2 \end{aligned}$$

$$+ \sum_k \left[\frac{V_{k,1}}{2} (1 + \cos \phi_k) + \frac{V_{k,2}}{2} (1 - \cos 2\phi_k) + \frac{V_{k,3}}{2} (1 + \cos 3\phi_k) \right] \quad (20)$$

Terms including the bond charges are explicitly shown here rather than being implicit, as they were before, in the terms including the MM point charges; however, both the MM point charges and bond charges are still implicitly in $H_{\mu\nu}^{\text{core}*}$. All variables and parameters are as just described.

CALCULATION OF GRADIENTS

If a geometry optimization is to be performed, then energy gradients with respect to the atomic coordinates must be calculated. Because there are contributions to the energy from pure QM terms (actually these terms include the electrostatic interactions between the MM partial charges and the QM electrons and nuclei), pure MM terms, and QM/MM terms, there are analogous contributions to the energy gradients with respect to the coordinate X_A :

$$\frac{\partial E_{\text{tot}}}{\partial X_A} = \frac{\partial E_{\text{QM}}}{\partial X_A} + \frac{\partial E_{\text{MM}}}{\partial X_A} + \frac{\partial E_{\text{QM/MM}}}{\partial X_A} \quad (21)$$

For the pure QM terms, the following serves as a starting point:

$$\begin{aligned} \frac{\partial E_{\text{QM}}^*}{\partial X_A} = & \sum_{\mu\nu} P_{\mu\nu} \frac{\partial H_{\mu\nu}^{\text{core}*}}{\partial X_A} \\ & + \frac{1}{2} \sum_{\mu\nu} \sum_{\alpha\beta} P_{\mu\nu} P_{\alpha\beta} \frac{\partial (\mu\nu || \alpha\beta)}{\partial X_A} \\ & + \sum_{CD} \frac{\partial}{\partial X_A} \left(\frac{Z_C Z_D}{R_{CD}} \right) + \sum_{CB} \frac{\partial}{\partial X_A} \left(\frac{Z_C q_B}{r_{CB}} \right) \\ & + \sum_{CM} \frac{\partial}{\partial X_A} \left(\frac{Z_C q_M}{r_{CM}} \right) + \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial X_A} H_{\mu\nu}^{\text{core}*} \\ & + \sum_{\mu\nu} \sum_{\alpha\beta} \frac{\partial P_{\mu\nu}}{\partial X_A} P_{\alpha\beta} (\mu\nu || \alpha\beta) \quad (22) \end{aligned}$$

where * indicates the inclusion of QM/MM electrostatic terms, C and D label QM nuclei, B labels bond charges, and M labels MM atoms. Calculation of the derivatives in the first four terms is relatively straightforward and proceeds in a similar manner as for regular HF calculations. The last

two terms, however, are far more complicated, with the remaining steps for the derivations of the gradient expression being detailed in the Appendix. The resulting expression is:

$$\begin{aligned}
 \frac{\partial E_{\text{QM}}^*}{\partial X_A} = & \sum_{\mu\nu} P_{\mu\nu} \frac{\partial H_{\mu\nu}^{\text{core}*}}{\partial X_A} \\
 & + \frac{1}{2} \sum_{\mu\nu} \sum_{\alpha\beta} P_{\mu\nu} P_{\alpha\beta} \frac{\partial(\mu\nu||\alpha\beta)}{\partial X_A} \\
 & + \sum_{CD} \frac{\partial}{\partial X_A} \left(\frac{Z_C Z_D}{R_{CD}} \right) + \sum_{CB} \frac{\partial}{\partial X_A} \left(\frac{Z_C q_B}{r_{CB}} \right) \\
 & + \sum_{CM} \frac{\partial}{\partial X_A} \left(\frac{Z_C q_M}{r_{CM}} \right) \\
 & + 4 \sum_{i=\text{fr}} \sum_{\mu\nu} \sum_{\alpha} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i} F_{\mu\nu}^* C_{\nu i} \\
 & - 4 \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} \epsilon_{ij}^* \frac{\partial R_{\gamma\alpha}^i}{\partial X_A} C_{\alpha i} S_{\gamma\delta} C_{\delta j} \\
 & - 2 \sum_{i,j=\text{fr}} \sum_{\gamma\delta} \epsilon_{ij}^* C_{\gamma i} \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j} \\
 & - 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} C_{\mu i} \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu j} \epsilon_{ij}^* \\
 & - 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} \sum_{\alpha} C_{\mu i} S_{\mu\nu} \frac{\partial R_{\nu\alpha}^j}{\partial X_A} C_{\alpha j} \epsilon_{ij}^* \\
 & - 4 \sum_{i=\text{non}} \sum_{g=\text{fr}} \sum_{\alpha\beta} C_{\alpha i} \frac{\partial S_{\alpha\beta}}{\partial X_A} C_{\beta g} \omega_{gi} \\
 & - 4 \sum_{i=\text{non}} \sum_{g=\text{fr}} \sum_{\alpha\beta\gamma} C_{\alpha i} S_{\alpha\beta} \frac{\partial R_{\beta\gamma}^g}{\partial X_A} C_{\gamma g} \omega_{gi}
 \end{aligned} \quad (23)$$

Here, ϵ_{ij}^* is defined as $\sum_{\mu\nu} C_{\mu i} F_{\mu\nu}^* C_{\nu j}$ and ω is defined as:

$$\omega_{gi} \equiv \sum_{j=\text{fr. virt.}} \sum_{f,h=\text{fr}} \sum_{\mu\nu} \mathbf{U}_{gh} \mathbf{w}_h^{-1} \mathbf{V}_{fh} \tilde{\mathbf{C}}_{\mu f} \mathbf{S}_{\mu\nu} C_{\nu j} \epsilon_{ji}^* \quad (24)$$

where $\tilde{\mathbf{C}}$ is the coefficient matrix for the frozen orbitals projected onto the basis functions allowed for use by the nonfrozen orbitals, and \mathbf{U} , \mathbf{w} , and \mathbf{V} are found from the singular value decomposition of $(\tilde{\mathbf{C}}^\dagger \mathbf{S} \mathbf{C}) = \mathbf{V} \mathbf{w} \mathbf{U}^\dagger$. How the derivatives for the rotation matrix elements (derivatives of the form $\partial R_{\mu\alpha}^i / \partial X_A$) are calculated is shown in the Appendix.

Calculations for the MM gradients are straightforward; expressions for such derivatives simply involve applying the chain rule to the relevant energy terms once these terms have been expressed in terms of the atomic coordinates. The QM/MM energy terms are of the same form as the bonding and nonbonding MM terms, and therefore the QM/MM gradient calculations are also straightforward and proceed in the same manner as for the MM terms.

It should be noted that the bond charges are set to be at the midpoint of the two bond-adjacent atoms and thus are not displaced independently of the two bond-adjacent atoms. Therefore, the gradients calculated with respect to the coordinates of the bond charges need to be related to the gradients with respect to the coordinates of the two bond-adjacent atoms. If the two bond-adjacent atoms (the first is the MM atom, whereas the second is the QM atom) are labeled M and Q , respectively, and the bond charge is labeled B , then the x -coordinates are related in the following way:

$$x_B = 0.5x_M + 0.5x_Q \quad (25)$$

Identical expressions exist for the y - and z -coordinates. Now, the total gradients on M and Q are:

$$\begin{aligned}
 \left(\frac{\partial E}{\partial x_M} \right)_{\text{tot}} &= \left(\frac{\partial E}{\partial x_M} \right) + \left(\frac{\partial E}{\partial x_B} \right) \left(\frac{\partial x_B}{\partial x_M} \right) \\
 &= \left(\frac{\partial E}{\partial x_M} \right) + 0.5 \left(\frac{\partial E}{\partial x_B} \right) \quad (26)
 \end{aligned}$$

$$\begin{aligned}
 \left(\frac{\partial E}{\partial x_Q} \right)_{\text{tot}} &= \left(\frac{\partial E}{\partial x_Q} \right) + \left(\frac{\partial E}{\partial x_B} \right) \left(\frac{\partial x_B}{\partial x_Q} \right) \\
 &= \left(\frac{\partial E}{\partial x_Q} \right) + 0.5 \left(\frac{\partial E}{\partial x_B} \right) \quad (27)
 \end{aligned}$$

with identical derivative expressions for the y - and z -coordinates. If a position between atoms M and Q other than the midpoint were chosen for the bond charge, then these equations would have to reflect that.

As a final note on the gradients, all of the analytical expressions just shown have been tested against numerical finite-difference results. The two sets of results for the gradients showed very favorable agreement.

Tests with Alanine Dipeptide and Tetrapeptide

The mixed *ab initio* QM/MM method using frozen orbitals has been tested by determining the optimized geometries and relative energies for ten conformations of alanine tetrapeptide. Alanine tetrapeptide consists of three central alanyl residues capped by an acetyl group on the N-terminal end and by an *N*-methylamine group on the C-terminal end. The ten conformations used are the same as those used by Beachy et al., with the full QM geometry optimizations and energy calculations performed as described in that study.³⁰

The first step in performing the QM/MM modeling is to partition the alanine tetrapeptides into QM and MM regions. Figure 3 shows how the tetrapeptides are partitioned, with the location of the frozen bond being indicated. Also, the atoms for which basis functions will be included in the frozen orbital are circled.

The next step is to form the frozen orbital using a model molecule. The model molecule chosen is the C5 conformer of alanine dipeptide. In generating the frozen orbital, the SCF procedure using JAGUAR is performed for the C5 conformer at the restricted HF level with a 6-31G** basis set. After convergence is obtained, the ground-state wave function is Boys-localized and the molecular or-

bital corresponding to the " σ -orbital" between N_1 and C_2 (see Fig. 4 for the partitioning and numbering of the C5 conformer model molecule) is extracted as the frozen orbital.

Once the frozen orbital has been found, associated correction terms need to be determined. To find the stretch and bend correction parameters, the six minimum conformations ($C5$, $C7_{eq}$, $C7_{ax}$, β_2 , α_L , and α') of alanine dipeptide were each perturbed. Looking at Figure 4, these perturbations include stretching and contracting the bond distances between atoms N_1-C_2 , N_1-C_5 , N_1-H_6 , C_2-C_7 , C_2-C_{12} , and C_2-H_{13} and widening and narrowing the bond angles defined by atoms $C_2-N_1-C_5$, $C_2-N_1-H_6$, $N_1-C_2-C_7$, $N_1-C_2-C_{12}$, and $N_1-C_2-H_{13}$. The actual changes in bond length used were 0.03, 0.02, 0.01, 0.00, -0.01, -0.02, and -0.03 Å from the optimized bond length for each stretch (only one bond length was varied at a time), while the actual changes in bond angle used were 3.6°, 2.4°, 1.2°, 0.0°, -1.2°, -2.4°, and -3.6° from the optimized bond angle for each bend (again, only one angle was varied at a time, and bond lengths and angles were not changed at the same time). After all QM/MM and pure QM energies had been obtained for all perturbed geometries, the stretch and bend correction parameters were obtained simultaneously for all six conformational minima by a linear least-squares fitting procedure, as described earlier. The resulting parameters are shown in Table I.

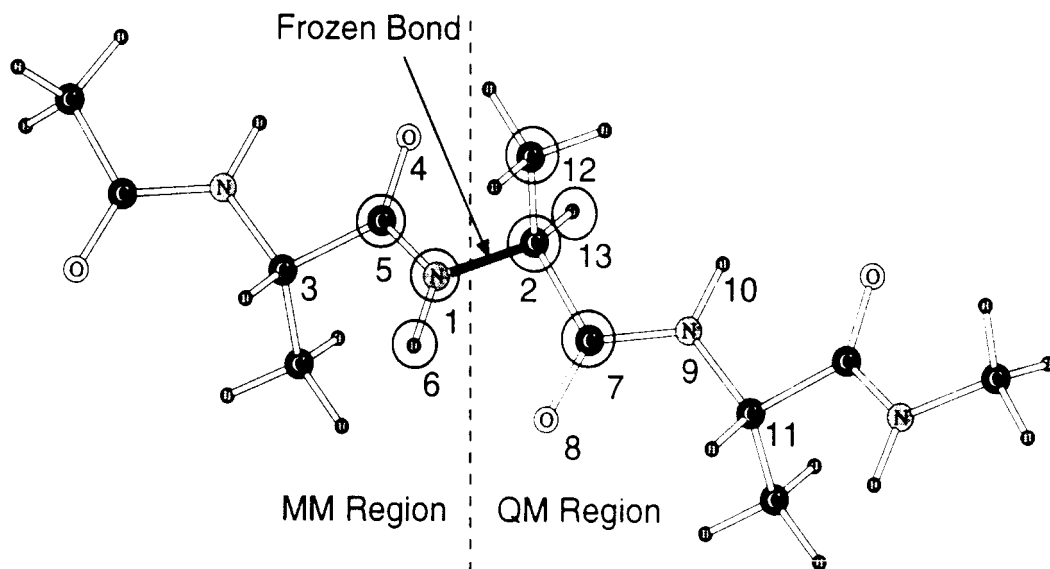


FIGURE 3. Partitioning of tetrapeptides into QM and MM regions. The location of the frozen bond is indicated, and the atoms for which basis functions will be included in the frozen orbital are circled.

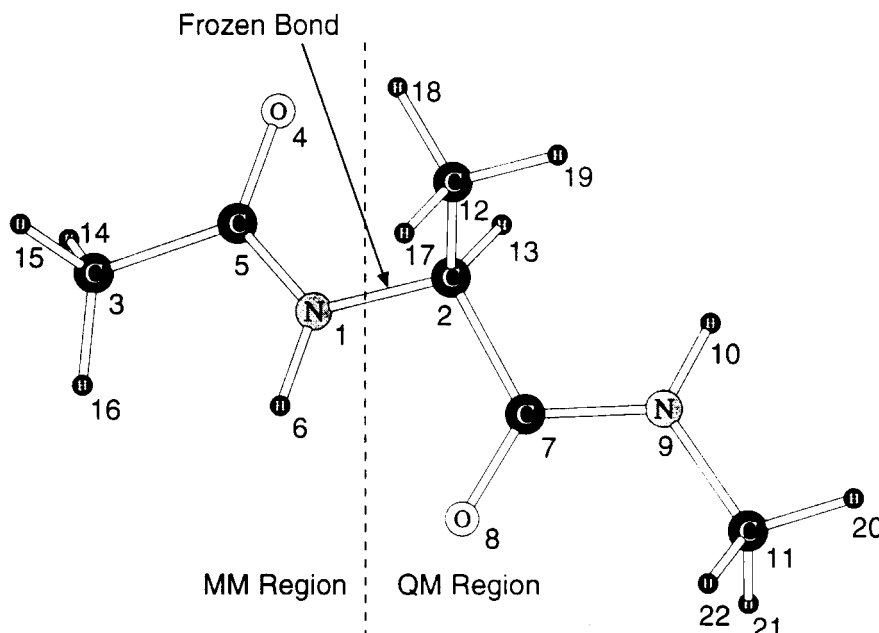


FIGURE 4. Partitioning and numbering for C5 conformer model molecule.

TABLE I.
Parameters for Stretch, Bend, and Torsion
Correction Terms.^a

Atoms	Stretches		
	k_i	r_{i0}	
N ₁ —C ₂	−396.4	1.717	
N ₁ —C ₅	−134.0	1.645	
N ₁ —H ₆	113.8	0.727	
C ₂ —C ₇	−76.3	1.361	
C ₂ —C ₁₂	−93.1	1.797	
C ₂ —H ₁₃	−35.7	1.053	
Atoms	Bends		
	k_j	θ_{j0}	
C ₂ —N ₁ —C ₅	−6.2	398.05	
C ₂ —N ₁ —H ₆	45.3	99.64	
N ₁ —C ₂ —C ₇	−37.5	125.53	
N ₁ —C ₂ —C ₁₂	−38.9	160.14	
N ₁ —C ₂ —H ₁₃	−66.9	121.04	
Atoms	Torsions		
	$V_{k,1}$	$V_{k,2}$	$V_{k,3}$
C ₅ —N ₁ —C ₂ —C ₇	2.680	−1.971	0.012
N ₁ —C ₂ —C ₇ —N ₉	1.543	2.591	−0.726

^a Units are kcal · mol^{−1} · Å^{−2} for k_i ; Å for r_{i0} ; kcal · mol^{−1} · degrees^{−2} for k_j ; degrees for θ_{j0} ; and kcal/mol for $V_{k,n}$.

The torsion correction parameters were determined by performing restricted geometry optimizations on the six minimum conformers and several other geometries where ϕ (the torsional angle defined by atoms C₅—N₁—C₂—C₇) and ψ (the torsional angle defined by atoms N₁—C₂—C₇—N₉) were set to specific values. For these restricted minimizations, the values for ϕ and ψ were kept fixed while the other degrees of freedom were allowed to vary. Also, the stretch and bend corrections found earlier were included in the energy calculations here. The resulting QM/MM and pure QM energies were compared, with the pure QM values in this case being the LMP2/cc-pVTZ(-f) energy values (obtained for the HF/6-31G** geometries). Here, LMP2 refers to localized Møller–Plesset second-order perturbation theory,^{31,32} whereas cc-pVTZ(-f) refers to the triple- ζ correlation-consistent basis of Dunning,³³ excluding f functions on second-row elements and d functions on hydrogen. The rationale for using the LMP2/cc-pVTZ(-f) values involves the fact that there is correlation energy involved between the QM and MM regions due to the presence of the van der Waals terms. The energy differences were then fit with a weighted linear least-squares fitting procedure, as described earlier.

To determine how well the QM/MM performed, mixed QM/MM unrestricted geometry optimizations were executed for the six conforma-

tional minima of alanine dipeptide (i.e., using these six conformers as the QM/MM molecules). The resulting relative conformational energies after these unrestricted minimizations are displayed in Table II, where they are also compared with LMP2/cc-pVTZ(-f), HF/6-31G**, and OPLS/IMPACT values. For validation purposes, absolute energies for the C5 conformer were: -283.278531 hartrees (QM/MM) and -492.884760 hartrees (HF/6-31G**). The ϕ and ψ angles for the resulting geometries are shown in Table III, where they are compared with the pure QM (HF/6-31G**) and OPLS/IMPACT values. The OPLS/IMPACT values were obtained by inputting the QM geometries as initial guesses and performing unrestricted optimizations employing the conjugate-gradient optimization method in IMPACT with convergence criteria of an RMS gradient < 0.001 kcal/(mol · Å) and a change in energy between iterations of < 0.0001 kcal/mol. It should be noted that the QM/MM and OPLS/IMPACT methods were not able to find the β_2 and α_L conformations (OPLS/IMPACT was also not able to find the α' conformer), as these methods are not able to model the potentials in these regions accurately enough to give these shallow minima as stable points. For the

minima the QM/MM modeling was able to locate, relatively good agreement was found for the relative energies and geometries. The relative energy for the C7_{ax} conformation was somewhat low which may be indicative of the hydrogen-bonding-type interactions between the QM and MM regions being too strong and/or the torsional profile not being accurate enough. Also, ψ for the C7_{ax} conformer was about 19° off, because the potential is quite flat around this conformation, and the LMP2/cc-pVTZ(-f) energies for the HF/6-31G** geometries actually lead to a lower energy at this value for ψ .
Now that the frozen orbital had been determined and the correction terms parameterized, the QM/MM modeling of the alanine tetrapeptides could begin. This first involved manipulating the frozen orbital so that they could be used in the tetrapeptides. All contributions to the frozen orbital from basis functions centered on atoms other than the circled atoms in Figure 3 were removed. The frozen orbital was then rotated to be aligned along the frozen bond axis (the N₁—C₂ axis shown in Fig. 3) and renormalized (in QM/MM molecules where more than one frozen orbital was used, this step would not only involve renormalization of the

TABLE II.
Comparison of Relative Alanine Dipeptide Conformational Energies.^a

Conf.	QM/MM	LMP2/cc-pVTZ(-f) ³⁰	HF/6-31G**	OPLS/IMPACT
C5	0.89	0.95	0.37	1.32
C7 _{eq}	0.00	0.00	0.00	0.00
C7 _{ax}	2.25	2.67	2.85	2.53
β_2	→ C7 _{eq}	2.75	2.54	→ C7 _{eq}
α_L	→ C7 _{ax}	4.31	4.74	→ C7 _{ax}
α'	5.48	5.51	5.81	→ C7 _{ax}

^a QM/MM values are for unrestricted minimizations that include all correction energy terms. All energies are in kcal/mol.

TABLE III.
Comparison of Torsional Angles for Alanine Dipeptide Conformers.^a

Conf.	QM/MM		QM		OPLS/IMPACT	
	ϕ	ψ	ϕ	ψ	ϕ	ψ
C5	-149.7	160.9	-157.9	160.3	-145.0	161.1
C7 _{eq}	-81.3	73.3	-85.8	78.5	-80.2	67.6
C7 _{ax}	69.6	-37.4	75.8	-56.5	68.4	-55.0
β_2	→ C7 _{eq}	→ C7 _{eq}	-128.6	23.2	→ C7 _{eq}	→ C7 _{eq}
α_L	→ C7 _{ax}	→ C7 _{ax}	66.9	29.7	→ C7 _{ax}	→ C7 _{ax}
α'	-155.3	-32.1	-166.4	-40.1	→ C7 _{ax}	→ C7 _{ax}

^a QM/MM values are after unrestricted minimizations with all corrections included, whereas the QM values refer to HF/6-31G**-optimized geometries. All angles are given in degrees.

TABLE IV.
Comparison of Relative Tetrapeptide Conformational Energies.^a

Conf. #	QM/MM	LMP2/cc-pVTZ(-f) ³⁰	HF/6-31G**	OPLS/IMPACT
1	2.97	2.71	0.94	3.05
2	3.53	2.84	1.53	2.69
3	-0.25	0.00	0.71	-1.14
4	4.27	4.13	3.09	3.83
5	2.90	3.88	3.12	4.80
6	-0.28	2.20	1.95	-0.69
7	4.25	5.77	7.07	4.36
8	6.23	4.16	5.05	6.88
9	8.47	6.92	7.94	6.30
10	7.52	6.99	8.20	9.52
RMS	1.30	0.00	1.10	1.64

^a Also shown are RMS energy deviations for all conformers relative to the LMP2/cc-pVTZ(-f) values. All energies are in kcal/mol.

conformers 8 and 9 may be too high due to correction terms (especially the electrostatic terms) not being optimal and/or due to a nonoptimal arrangement of charge at the QM/MM interface (i.e., the representation of the charge that the MM region sees as the frozen orbital and charges of -0.2 and 0.9, respectively, placed on atom N₁ and at the bond midpoint between N₁ and C₂). These last potential difficulties could affect the rest of the energies as well. However, it must be kept in mind throughout this analysis that the QM/MM method is a hybrid of the pure QM and pure MM methods, and thus it is difficult to determine the exact origin of these energy discrepancies and, furthermore, it is not possible to obtain better results for the MM region and QM/MM interface than the MM results themselves.

As for the QM/MM geometries for the ten alanine tetrapeptide conformations, the RMSD values for the heavy atoms (nonhydrogen atoms) are given in Table V. The RMSD values are for derivations from the QM (HF/6-31G**) geometries, and are compared with analogous values obtained for the pure MM (OPLS/IMPACT) geometries. The ϕ and ψ torsional angles for the ten conformational structures after QM/MM optimization are given in Table VI and are compared with the pure QM and pure MM values. As can be seen from these two tables, the QM/MM geometries compare favorably to the pure QM structures and are again definitely no worse than the pure MM geometries. The QM/MM geometries for conformers 5, 6, and 8 had somewhat high RMSD values, whereas, their ψ_2 values were a fair amount different from the pure QM values. As discussed earlier in regard to

TABLE V.
Comparison of RMSD Values for Tetrapeptide Conformers.^a

Conf. #	QM/MM	OPLS/IMPACT
1	0.11	0.25
2	0.14	0.22
3	0.28	0.28
4	0.13	0.17
5	0.30	0.21
6	0.50	0.69
7	0.21	0.27
8	0.41	0.46
9	0.16	0.12
10	0.19	0.17

^a RMSD values are in Å and are expressed for only the heavy (nonhydrogen) atoms. Note that the LMP2/cc-pVTZ(-f) and HF/6-31G** RMSD values are 0.0

relative energies, reasons for these discrepancies could be the torsional profile not being reproduced accurately enough by the torsional corrections, the electrostatic corrections not being optimal, the charge representation at the QM/MM interface not yielding an accurate enough potential, and/or the hydrogen-bonding-type interactions between the QM and MM regions being too strong. It should also be noted that the QM/MM geometries for the QM regions beyond the QM/MM interface of the tetrapeptides, as expected, agree very well with those calculated by pure QM.

As stated earlier, the relative energies and geometries obtained for the alanine tetrapeptide conformers through QM/MM modeling agree favorably with the pure QM values and are definitely

TABLE VI.
Comparison of Torsional Angles for Tetrapeptide Conformers.^a

Conf. #	ϕ_1	ψ_1	ϕ_2	ψ_2	ϕ_3	ψ_3
1	-146.6	162.8	-149.2	167.1	-156.8	160.8
	(-158.5)	(163.5)	(-157.8)	(163.4)	(-156.2)	(160.8)
	[-146.6]	[164.4]	[-143.6]	[165.9]	[-142.9]	[163.2]
2	-147.2	164.9	-145.5	164.0	-86.1	79.2
	(-158.6)	(163.9)	(-154.9)	(158.1)	(-86.0)	(79.2)
	[-145.5]	[163.6]	[-141.3]	[162.7]	[-80.8]	[70.3]
3	-79.2	76.6	72.0	-50.5	-79.9	86.9
	(-81.7)	(93.4)	(76.3)	(-53.4)	(-80.5)	(85.1)
	[-78.6]	[75.7]	[71.9]	[-49.9]	[-74.0]	[76.3]
4	-143.3	158.7	-85.1	94.7	-160.3	150.3
	(-156.9)	(161.3)	(-88.8)	(83.5)	(-156.0)	(152.8)
	[-145.0]	[162.0]	[-82.8]	[74.1]	[-135.9]	[159.3]
5	-140.4	159.9	-88.3	0.8	-154.8	161.9
	(-157.2)	(170.0)	(-76.2)	(-19.6)	(-153.8)	(160.8)
	[-141.9]	[159.5]	[-77.1]	[-24.9]	[-144.8]	[157.6]
6	-88.6	60.0	71.2	-1.8	-166.5	152.4
	(-89.0)	(67.3)	(63.0)	(24.3)	(-165.0)	(149.8)
	[-89.3]	[63.6]	[68.5]	[-53.8]	[-89.3]	[142.3]
7	58.4	-155.7	-84.8	64.2	-161.6	-53.3
	(56.0)	(-158.5)	(-93.0)	(63.8)	(-163.3)	(-50.0)
	[56.5]	[-149.3]	[-87.0]	[57.1]	[-155.4]	[-56.9]
8	70.0	-62.5	-71.5	116.3	60.6	31.0
	(72.8)	(-70.5)	(-58.1)	(134.7)	(62.0)	(25.7)
	[68.8]	[-54.1]	[-62.5]	[98.3]	[64.1]	[33.0]
9	68.3	-51.1	69.7	-45.3	75.1	-52.9
	(75.7)	(-59.5)	(76.1)	(-55.3)	(75.5)	(-53.0)
	[67.7]	[-53.8]	[67.2]	[-52.3]	[67.4]	[-52.7]
10	47.7	30.1	73.8	12.3	75.2	-51.5
	(62.5)	(29.0)	(65.1)	(20.6)	(73.8)	(-51.5)
	[58.9]	[-2.7]	[56.6]	[44.7]	[71.7]	[-51.1]

^a Values for ϕ_i and ψ_i are for the QM/MM geometries with the analogous values for the pure QM (HF/6-31G**) geometries given in parentheses and for the pure MM (OPLX/IMPACT) geometries given in brackets under the QM/MM values. All angles are given in degrees, and numbering begins from the N-terminal end.

no worse than the pure MM values. However, some discrepancies are prevalent, suggesting that some improvements could be attempted for QM/MM modeling. As hinted earlier, some possibilities for improvement include modifying the electrostatic correction terms and/or the charge representation at the QM/MM interface. For the latter, the bond charge and charge on the MM bond-adjacent atom could be adjusted or a different charge arrangement at the QM/MM interface could be used. This could involve placing charges along the bonds between atoms N_1-C_5 and N_1-H_6 in addition to the bond charge between N_1-C_2 and the charge on N_1 (the positions of the

charges could be varied as well as the charge themselves) or not letting the MM region experience the frozen orbital but rather just split up the frozen orbital into point charges placed on N_1 , C_2 , and probably somewhere along the frozen bond N_1-C_2 (this last charge would probably be necessary to maintain the proper dipole moment acting on the MM region). Other possibilities for improvement could involve doing something to weaken the hydrogen-bonding-type interactions between the QM and MM regions (at least for shorter range) and sampling more geometries and/or using a better weighting scheme (one more directly dependent on the torsional gradients so

that the points around the minima are more highly weighted) when performing parameterization of the torsional corrections.

Conclusion

The methodology for mixed *ab initio* QM/MM modeling using frozen orbitals has been presented. A frozen orbital with associated correction parameters was formulated, using the alanine dipeptide conformational minima, for utilization at one of the locations where peptides and proteins could be "cut" to be partitioned into QM and MM regions. The method was then tested on ten conformers of alanine tetrapeptide to calculate conformational geometries and relative energies, as well as on the six dipeptide minima used in determining the frozen orbital and associated corrections. The QM/MM results compared favorably to those obtained from pure QM calculations and were definitely no worse than the pure MM results. This is important for a QM/MM method, in which the object is required to be able to perform rigorous calculations on a small region of special interest without unreasonable energy and structural errors resulting from the modeling of the MM region and the QM/MM interface.

Although the current QM/MM methodology may already be adequate for utilization in real applications, it is beneficial to explore possible improvements, such as those discussed in this work. Once this has been completed, a database of frozen orbitals and associated parameters will need to be built before commencement of QM/MM modeling for real applications such as the active sites of enzymes. Also, the addition of electron correlation will probably be needed, and modification of the method should be possible to readily incorporate such techniques as LMP2 or density functional theory (DFT).^{34–37} Modeling of other systems, including material systems such as semiconductor surfaces and catalysts, should then be viable.

Appendix

DETERMINATION OF ROTATION MATRICES FOR FROZEN ORBITALS

The first step in finding the matrices for rotating the frozen orbitals is to find the 3×3 matrix that transforms the reference frame of the model

molecule to that of the QM/MM molecule. In Figure 2, the reference frame for the model molecule was defined by $C_3-N_1-C_2$, whereas the reference frame for the QM/MM molecule was defined by $C'_3-N'_1-C'_2$. These reference frames are reproduced in Figure 6, where vectors that will be used in subsequent calculations are defined.

The first entities calculated are the matrices to transform vectors \mathbf{d} , \mathbf{c} , and \mathbf{a} of the model molecule to be aligned, in the \hat{x} , \hat{y} , and \hat{z} directions (matrix \mathbf{P}_1), respectively, and to transform vectors \mathbf{d}' , \mathbf{c}' , and \mathbf{a}' of the QM/MM molecule to be aligned in the \hat{x} , \hat{y} , and \hat{z} directions (matrix \mathbf{P}_2^{-1}), respectively. These matrices are found as the directional cosine matrices:

$$\mathbf{P}_1 = \begin{pmatrix} \frac{d_x}{d} & \frac{d_y}{d} & \frac{d_z}{d} \\ \frac{c_x}{c} & \frac{c_y}{c} & \frac{c_z}{c} \\ \frac{a_x}{a} & \frac{a_y}{a} & \frac{a_z}{a} \end{pmatrix}; \quad (28)$$

$$\mathbf{P}_2^{-1} = \begin{pmatrix} \frac{d'_x}{d'} & \frac{d'_y}{d'} & \frac{d'_z}{d'} \\ \frac{c'_x}{c'} & \frac{c'_y}{c'} & \frac{c'_z}{c'} \\ \frac{a'_x}{a'} & \frac{a'_y}{a'} & \frac{a'_z}{a'} \end{pmatrix}$$

Once \mathbf{P}_1 and \mathbf{P}_2 have been found, the rotation matrix that aligns the reference frame of the model molecule with that of the QM/MM molecule is:

$$\mathbf{P}_{\text{tot}} = \mathbf{P}_2 \mathbf{P}_1 \quad (29)$$

\mathbf{P}_{tot} will be used to determine the part of the orbital rotation matrix that rotates basis functions

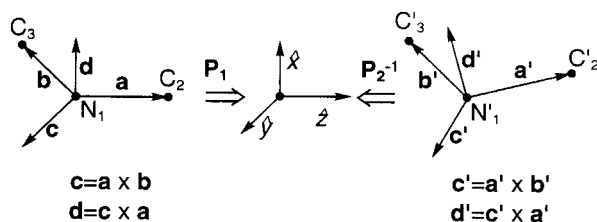


FIGURE 6. Reference frames for rotations with definition of vectors used in subsequent calculations. Note that \mathbf{d} is the projection of \mathbf{b} onto the plane perpendicular to the frozen bond axis \mathbf{a} (\mathbf{d}' is the projection of \mathbf{b}' onto the plane perpendicular to \mathbf{a}'). In the subsequent calculations, $\mathbf{a} = a_x \hat{x} + a_y \hat{y} + a_z \hat{z}$, $a = |\mathbf{a}|$, etc.

on atoms N_1 , C_2 , and C_3 (it will be the actual matrix that rotates the p basis functions), as shown in what follows. For the other atoms i connected to the frozen bond-adjacent atoms, the additional azimuthal rotations by ϕ_i must be taken into account. Figure 2c demonstrates how ϕ is determined for each atom (note that both the model molecule and QM/MM molecule reference frames have been aligned along the \hat{x} , \hat{y} , and \hat{z} directions for the determination of ϕ). Once ϕ_i has been found for each of the connected atoms i , the additional torsional matrix:

$$\mathbf{P}_{\text{tor}}(\phi_i) = \begin{pmatrix} \cos(\phi_i) & -\sin(\phi_i) & 0 \\ \sin(\phi_i) & \cos(\phi_i) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (30)$$

is calculated for each of these atoms. $\mathbf{P}_{\text{tor}}(\phi_i)$ is then used in conjunction with the \mathbf{P}_1 and \mathbf{P}_2 matrices found previously to get the total rotation matrix for each of these atoms:

$$\mathbf{P}_{\text{tot}}^i = \mathbf{P}_2 \mathbf{P}_{\text{tor}}(\phi_i) \mathbf{P}_1 \quad (31)$$

$$\mathbf{D}^i = \begin{pmatrix} P_{1,1}^2 & P_{1,2}^2 & P_{1,3}^2 & 2\sqrt{3} P_{1,1} P_{1,2} & 2\sqrt{3} P_{1,1} P_{1,3} & 2\sqrt{3} P_{1,2} P_{1,3} \\ P_{2,1}^2 & P_{2,2}^2 & P_{2,3}^2 & 2\sqrt{3} P_{2,1} P_{2,2} & 2\sqrt{3} P_{2,1} P_{2,3} & 2\sqrt{3} P_{2,2} P_{2,3} \\ P_{3,1}^2 & P_{3,2}^2 & P_{3,3}^2 & 2\sqrt{3} P_{3,1} P_{3,2} & 2\sqrt{3} P_{3,1} P_{3,3} & 2\sqrt{3} P_{3,2} P_{3,3} \\ \frac{1}{\sqrt{3}} P_{1,1} P_{2,1} & \frac{1}{\sqrt{3}} P_{1,2} P_{2,2} & \frac{1}{\sqrt{3}} P_{1,2} P_{2,3} & P_{1,1} P_{2,2} + P_{1,2} P_{2,1} & P_{1,1} P_{2,3} + P_{1,3} P_{2,1} & P_{1,2} P_{2,3} + P_{1,3} P_{2,2} \\ \frac{1}{\sqrt{3}} P_{1,1} P_{3,1} & \frac{1}{\sqrt{3}} P_{1,2} P_{3,2} & \frac{1}{\sqrt{3}} P_{1,3} P_{3,3} & P_{1,1} P_{3,2} + P_{1,2} P_{3,1} & P_{1,1} P_{3,3} + P_{1,3} P_{3,1} & P_{1,2} P_{3,3} + P_{1,3} P_{3,2} \\ \frac{1}{\sqrt{3}} P_{2,1} P_{3,1} & \frac{1}{\sqrt{3}} P_{2,2} P_{3,2} & \frac{1}{\sqrt{3}} P_{2,3} P_{3,3} & P_{2,1} P_{3,2} + P_{2,2} P_{3,1} & P_{2,1} P_{3,3} + P_{2,3} P_{3,1} & P_{2,2} P_{3,3} + P_{2,3} P_{3,2} \end{pmatrix} \quad (34)$$

where it is assumed that the d functions are $(d_{x^2}$, d_{y^2} , d_{z^2} , d_{xy} , d_{xz} , d_{yz}), respectively. The factors of $\sqrt{3}$ are normalization factors.

Finally, the orbital rotation matrix \mathbf{R} is constructed. First, it must be made clear that only basis functions on the same atom and of the same angular momentum and principle quantum number can participate in the same transformation (e.g., the $3d$ functions for the N_1 atom are only rotated among themselves). Thus, \mathbf{R} will be in block diagonal form, with each block being the rotation matrix for a particular block of basis functions. Going back to the example presented in Figure 2, and

Note that $\mathbf{P}_{\text{tot}}^i$ is only equal to $\mathbf{P}_{\text{tot}} = \mathbf{P}_2 \mathbf{P}_1$ when i is the reference atom or one of the atoms adjacent to the frozen bond.

To proceed with the rotations of basis functions, it is first noted that the s functions are spherically symmetric and thus invariant to rotations. The matrix to rotate s functions is:

$$\mathbf{S} = (S_{1,1}) = (1) \quad (32)$$

The p functions transform like x , y , and z , and therefore the matrix to rotate p functions is simply the matrix to rotate an atom i in the frame of the model molecule to the frame of the QM/MM molecule:

$$\mathbf{P}^i = \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{pmatrix} = \mathbf{P}_{\text{tot}}^i \quad (33)$$

Matrices for rotations of basis functions with angular momentum greater than p functions, such as d, f, g, \dots , can be derived from \mathbf{P} . Only the matrix for d functions, however, is presented here:

assuming that the basis set is 6-31G**, the order of basis functions is written as:

$$\left\{ \phi_{1s}(N_1), \phi_{2s}(N_1), \phi_{2p_x}(N_1), \dots, \right. \\ \left. \phi_{3d_{yz}}(N_1), \phi_{1s}(C_2), \dots, \right. \\ \left. \phi_{3d_{yz}}(C_6), \phi_{1s}(H_7), \phi_{2s}(H_7), \right. \\ \left. \phi_{2p_x}(H_7), \phi_{2p_y}(H_7), \phi_{2p_z}(H_7) \right\}.$$

There will be other basis functions centered on other atoms, but none of these enter into this particular frozen orbital. The orbital rotation ma-

trix \mathbf{R} is then written as:

$$\mathbf{R} = \begin{pmatrix} \begin{matrix} \boxed{\mathbf{S}} & & & & & \\ & \boxed{\mathbf{S}} & & & & \\ & & \boxed{\mathbf{P}} & & & \\ & & & \boxed{\mathbf{S}} & & \\ & & & & \boxed{\mathbf{P}} & \\ & & & & & \boxed{\mathbf{D}} \end{matrix} & \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \bigg\} \begin{matrix} \mathbf{N}_1 \\ \\ \\ \\ \\ \end{matrix} & \mathbf{0} \\ \mathbf{0} & & \mathbf{H}_7 \begin{matrix} \boxed{\mathbf{S}} \\ \boxed{\mathbf{S}} \\ \boxed{\mathbf{P}} \end{matrix} \end{pmatrix}$$

where the first six blocks shown are the blocks for rotating the $1s$, $2s$, $2p$, $3s$, $3p$, and $3d$ basis functions centered on the N_1 atom, and the last three blocks are for the $1s$, $2s$, and $2p$ functions on the H_7 atom, respectively. All \mathbf{S} blocks for the N_1 functions are simply (1), whereas the \mathbf{P} and \mathbf{D} blocks for N_1 are derived as illustrated above, with the two \mathbf{P} blocks being the same 3×3 matrix. The \mathbf{S} blocks for H_7 are likewise simply (1), whereas the \mathbf{P} block is the specially determined matrix with the inclusion of the $\mathbf{P}_{\text{tor}}(\phi_i)$ matrix for the H_7 atom, as just demonstrated.

As mentioned earlier in this work, a separate orbital rotation matrix, \mathbf{R}^a , is computed for each frozen orbital a . Thus, when rotating all the frozen orbitals (represented by their coefficient matrix \mathbf{C}), the rotation matrices form a tensor of rank three, \mathcal{R} .

DERIVATION OF MODIFIED ROOHTAAN EQUATIONS

To derive the modified Roothaan equations in the presence of frozen orbitals, functional variation is used. The starting point is the expectation value for the ground-state energy of the restricted closed-shell single determinant wave function $|\Psi_0\rangle = |\psi_1\bar{\psi}_1 \cdots \psi_i\bar{\psi}_i \cdots \psi_{N/2}\bar{\psi}_{N/2}\rangle$, where $N/2$ is the total number of doubly occupied (frozen + nonfrozen) orbitals. The expectation value, $E_0[\{\psi_i\}]$, is a functional of the spatial orbitals $\{\psi_i\}$ and is given as:

$$E_0 = \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle = 2 \sum_i (i|h|i) + \sum_i \sum_j 2(i\bar{i}|j\bar{j}) - (ij|ji) \quad (36)$$

Expanding the spatial orbitals in a basis gives:

$$E_0 = 2 \sum_i \sum_{\mu\nu} C_{\mu i} C_{\nu i} (\mu|h|\nu) + \sum_i \sum_j \sum_{\mu\nu} \sum_{\alpha\beta} C_{\mu i} C_{\nu i} C_{\alpha j} C_{\beta j} \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \quad (37)$$

Now, using the orthonormality constraints:

$$\mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{1}, \text{ which is equivalent to } \sum_{\mu\nu} C_{\mu i} S_{\mu\nu} C_{\nu j} = \delta_{ij} \quad (38)$$

the functional, \mathcal{L} , is defined as:

$$\mathcal{L} = E_0 - 2 \sum_i \sum_j \epsilon_{ij} \left[\sum_{\mu\nu} C_{\mu i} S_{\mu\nu} C_{\nu j} - \delta_{ij} \right] \quad (39)$$

where ϵ_{ij} are undetermined Lagrange multipliers. The spatial orbitals (meaning the expansion coefficients, $C_{\mu i}$) are now varied to minimize the functional \mathcal{L} and thus E_0 . Because the frozen orbitals remain frozen during the SCF calculations, variations in their expansion coefficients are 0, and therefore they do not enter in the following variational equation:

$$\begin{aligned} \delta \mathcal{L} = & 2 \sum_{i=\text{non}} \sum_{\mu\nu} [\delta C_{\mu i} H_{\mu\nu}^{\text{core}} C_{\nu i} + C_{\mu i} H_{\mu\nu}^{\text{core}} \delta C_{\nu i}] \\ & + \sum_{i=\text{non}} \sum_j \sum_{\mu\nu} \sum_{\alpha\beta} \delta C_{\mu i} C_{\nu i} C_{\alpha j} C_{\beta j} \\ & \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \\ & + \sum_{i=\text{non}} \sum_j \sum_{\mu\nu} \sum_{\alpha\beta} C_{\mu i} \delta C_{\nu i} C_{\alpha j} C_{\beta j} \\ & \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \\ & + \sum_{j=\text{non}} \sum_i \sum_{\mu\nu} \sum_{\alpha\beta} C_{\mu i} C_{\nu i} \delta C_{\alpha j} C_{\beta j} \\ & \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \\ & + \sum_{j=\text{non}} \sum_i \sum_{\mu\nu} \sum_{\alpha\beta} C_{\mu i} C_{\nu i} C_{\alpha j} \delta C_{\beta j} \\ & \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \\ & - 2 \sum_{i=\text{non}} \sum_j \sum_{\mu\nu} \epsilon_{ij} [\delta C_{\mu i} S_{\mu\nu} C_{\nu j} + C_{\mu i} S_{\mu\nu} \delta C_{\nu j}] \end{aligned} \quad (40)$$

where $H_{\mu\nu}^{\text{core}} = (\mu|h|\nu)$. Next, switching relevant indices and combining terms, the following is ob-

tained:

$$\begin{aligned}\delta\mathcal{L} = & 4 \sum_{i=\text{non}} \sum_{\mu\nu} \delta C_{\mu i} H_{\mu\nu}^{\text{core}} C_{\nu i} \\ & + 4 \sum_{i=\text{non}} \sum_j \sum_{\mu\nu} \sum_{\alpha\beta} \delta C_{\mu i} C_{\nu i} C_{\alpha j} C_{\beta j} \\ & \times [2(\mu\nu|\alpha\beta) - (\mu\alpha|\beta\nu)] \\ & - 4 \sum_{i=\text{non}} \sum_j \epsilon_{ij} \sum_{\mu\nu} \delta C_{\mu i} S_{\mu\nu} C_{\nu j} \quad (41)\end{aligned}$$

Using the following relations:

$$J_{\mu\nu} = \sum_j \sum_{\alpha\beta} C_{\alpha j} C_{\beta j} (\mu\nu|\alpha\beta) \quad (42)$$

$$K_{\mu\nu} = \sum_j \sum_{\alpha\beta} C_{\alpha j} C_{\beta j} (\mu\alpha|\beta\nu) \quad (43)$$

$\delta\mathcal{L}$ is set equal to 0 to find the minimum:

$$\begin{aligned}0 = \delta\mathcal{L} = & 4 \sum_{i=\text{non}} \sum_{\mu\nu} \delta C_{\mu i} [H_{\mu\nu}^{\text{core}} + 2J_{\mu\nu} - K_{\mu\nu}] C_{\nu i} \\ & - 4 \sum_{i=\text{non}} \sum_j \epsilon_{ij} \sum_{\mu\nu} \delta C_{\mu i} S_{\mu\nu} C_{\nu j} \quad (44)\end{aligned}$$

Using the relation, $F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + 2J_{\mu\nu} - K_{\mu\nu}$, the following simplified result is obtained:

$$\begin{aligned}0 = & \sum_{i=\text{non}} \sum_{\mu\nu} \delta C_{\mu i} F_{\mu\nu} C_{\nu i} \\ & - \sum_{i=\text{non}} \sum_j \epsilon_{ij} \sum_{\mu\nu} \delta C_{\mu i} S_{\mu\nu} C_{\nu j} \quad (45)\end{aligned}$$

The next step is to perform unitary transformations to “canonical” orbitals (actually, only the nonfrozen orbitals become canonical; the frozen orbitals remain as they are):

$$C' = CU; \quad U = \begin{array}{cc} & \begin{array}{cc} \text{fr} & \text{non} \end{array} \\ \begin{array}{c} \text{fr} \\ \text{non} \end{array} & \left(\begin{array}{c|c} \mathbf{1} & \mathbf{0} \\ \hline \mathbf{0} & \vdots \end{array} \right) \quad (46)$$

$$e' = U^\dagger \epsilon U; \quad \epsilon' = \left(\begin{array}{c|c} \epsilon_f & \epsilon'_{fn} \\ \hline \epsilon'_{nf} & \begin{array}{c} 0 \\ \cdot \\ \cdot \\ 0 \end{array} \end{array} \right). \quad (47)$$

Above, the unitary transformation matrix, U , consists of a unit matrix for the frozen orbital/frozen orbital block and zero matrices for the frozen/nonfrozen blocks, with the actual values of the nonfrozen/nonfrozen block being unimportant. What is important is that they, in principle, could be found such that the resulting matrix, ϵ' , consists of

a diagonal matrix for the nonfrozen/nonfrozen block. Because U is such that the frozen orbitals will remain unchanged, the frozen/nonfrozen blocks (ϵ'_{nf} and ϵ'_{fn}) of ϵ' will not be zero matrices; the frozen/frozen block (ϵ_f) will remain the same as in ϵ .

Once ϵ has been transformed to the canonical form of ϵ' (note that, in the following, the primes, when referring to elements of ϵ' , will be dropped for convenience), eq. (45) can be rewritten as:

$$\begin{aligned}0 = & \sum_{i=\text{non}} \sum_{\mu} \delta C_{\mu i} \\ & \times \sum_{\nu} \left[F_{\mu\nu} C_{\nu i} - \epsilon_i S_{\mu\nu} C_{\nu i} - \sum_{j=\text{fr}} \epsilon_{ij} S_{\mu\nu} C_{\nu j} \right] \quad (48)\end{aligned}$$

Because all the variations in the (nonfrozen) orbital coefficients are independent, eq. (48) is satisfied when the term multiplying each $\delta C_{\mu i}$ is set equal to 0:

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} - \sum_{\nu} \epsilon_i S_{\mu\nu} C_{\nu i} - \sum_{\nu} \sum_{j=\text{fr}} \epsilon_{ij} S_{\mu\nu} C_{\nu j} = 0 \quad (49)$$

Making the following definition:

$$F''_{\mu\nu} = F_{\mu\nu} - \sum_{i=\text{non}} \sum_{\alpha\beta} S_{\mu\alpha} C_{\alpha i} \epsilon_{ij} C_{\beta j} S_{\beta\nu} \quad (50)$$

the resulting modified Roothaan equation is:

$$\sum_{\nu} F''_{\mu\nu} C_{\nu i} = \sum_{\nu} \epsilon_i S_{\mu\nu} C_{\nu i}, \text{ or } \mathbf{F}'' \mathbf{C} = \mathbf{S} \mathbf{C} \epsilon \quad (51)$$

As mentioned earlier in this study, these equations are modified further when the nonfrozen orbitals are restricted to certain allowed basis functions. In this case, the initial nonfrozen occupied orbitals are made to be orthonormal to the frozen orbitals in the absence of the forbidden basis functions. Then, without the use of the forbidden basis functions, as many “nonfrozen virtuals” are made, such that the subspace of the allowed basis functions is spanned completely. The remaining virtual orbitals necessary to complete the basis spanning the space of all the basis functions are formed and are referred to as “frozen virtuals.” These frozen virtuals are made to be eigenvectors of the Fock matrix are thus prevented from mixing with the nonfrozen orbitals, in the same way as the frozen orbitals were, thus keeping the nonfrozen orbitals from being contaminated by the forbidden basis functions. Because the frozen virtuals are treated in the same way as the frozen orbitals (of course

they do not enter into the ground-state energy), the derivation of the Roothaan equation proceeds in the same way as before. The result is that there is now an additional term in the final equation [eq. (49)]:

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} - \sum_{\nu} \epsilon_i S_{\mu\nu} C_{\nu i} - \sum_{\nu} \sum_{j=\text{fr}} \epsilon_{ij} S_{\mu\nu} C_{\nu j} - \sum_{\nu} \sum_{j=\text{fr, virt.}} \epsilon_{ij} S_{\mu\nu} C_{\nu j} = 0 \quad (52)$$

which yields eqs. (7) and (8).

DERIVATION OF QM ENERGY GRADIENTS

The expression for the QM energy gradients starts with eq. (22). Calculations involving the first four terms are straightforward, whereas the last two need to be manipulated first. The first step is to substitute for the density \mathbf{P} in these last two terms, expand them, and split up the summation over molecular orbitals a into frozen (fr) and non-frozen (non) orbitals to obtain:

$$\begin{aligned} \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial X_A} H_{\mu\nu}^{\text{core}*} + \sum_{\mu\nu\alpha\beta} \frac{\partial P_{\mu\nu}}{\partial X_A} P_{\alpha\beta} (\mu\nu||\alpha\beta) \\ = 4 \sum_{\mu\nu} \sum_{a=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} F_{\mu\nu}^* C_{\nu a} \\ + 4 \sum_{\mu\nu} \sum_{a=\text{non}} \frac{\partial C_{\mu a}}{\partial X_A} F_{\mu\nu}^* C_{\nu a} \end{aligned} \quad (53)$$

Note that \mathbf{F}^* , the Fock energy matrix, and not \mathbf{F}^{*} , the matrix that satisfies the modified Roothaan equation, appears in eq. (53). It is the fact that \mathbf{F}^* no longer satisfies the Roothaan equation that makes these derivations so complex.

Now, the derivatives in the first term on the right-hand side of eq. (53) can be calculated explicitly for frozen orbitals a using eq. (4), which can be rewritten as:

$$C_{\mu a} = \sum_{\alpha} \sum_{i=\text{fr}} R_{\mu\alpha}^i C_{\alpha i}^0 U_{ia} \sigma_a \quad (54)$$

Here the notation \mathbf{C}^0 signifies that this is the (constant) coefficient matrix before the movement of atomic coordinates that would yield a new matrix \mathbf{C} , and the superscript i of $R_{\mu\alpha}^i$ indicates elements of the rotation tensor \mathcal{R} that rotate molecular orbital i . The elements of the frozen

density matrix can now be written as:

$$\begin{aligned} P_{\mu\nu} &= \sum_{a=\text{fr}} C_{\mu a} C_{\nu a} \\ &= \sum_{a=\text{fr}} \sum_{\alpha\beta} \sum_{i=\text{fr}} R_{\mu\alpha}^i C_{\alpha i}^0 U_{ia} \sigma_a^2 U_{aj} C_{\beta j}^0 R_{\nu\beta}^j \end{aligned} \quad (55)$$

Taking the derivatives of the elements of the frozen density matrix with respect to the atomic coordinate, X_A , yields:

$$\begin{aligned} \frac{\partial P_{\mu\nu}}{\partial X_A} &= \sum_{a=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} C_{\nu a}^0 + \sum_{a=\text{fr}} C_{\mu a}^0 \frac{\partial C_{\nu a}}{\partial X_A} \\ &= \sum_{\alpha} \sum_{i=\text{fr}} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i}^0 C_{\nu i}^0 \\ &\quad + \sum_{i=\text{fr}} \sum_{\alpha} C_{\mu i}^0 C_{\alpha i}^0 \frac{\partial R_{\nu\alpha}^i}{\partial X_A} \\ &\quad + \sum_{i,j,k=\text{fr}} C_{\mu i}^0 \frac{\partial U_{ij}}{\partial X_A} U_{jk} C_{\nu k}^0 \\ &\quad + \sum_{i,j,k=\text{fr}} C_{\mu i}^0 U_{ij} \frac{\partial U_{jk}}{\partial X_A} C_{\nu k}^0 \\ &\quad + \sum_{i,j,k=\text{fr}} C_{\mu k}^0 U_{ki} \frac{\partial (\sigma_i^2)}{\partial X_A} U_{il} C_{\nu l}^0 \end{aligned} \quad (56)$$

The third and fourth terms in the last equation can be combined as:

$$\begin{aligned} \sum_{i,j,k=\text{fr}} C_{\mu i}^0 \frac{\partial U_{ij}}{\partial X_A} U_{jk} C_{\nu k}^0 + \sum_{i,j,k=\text{fr}} C_{\mu i}^0 U_{ij} \frac{\partial U_{jk}}{\partial X_A} C_{\nu k}^0 \\ = \sum_{i,j,k=\text{fr}} C_{\mu i}^0 \left(\frac{\partial U_{ij}}{\partial X_A} U_{jk} + U_{ij} \frac{\partial U_{jk}}{\partial X_A} \right) C_{\nu k}^0 = 0 \end{aligned} \quad (58)$$

due to the fact that \mathbf{U} is unitary and thus

$$\sum_j \left(\frac{\partial U_{ij}}{\partial X_A} U_{jk} + U_{ij} \frac{\partial U_{jk}}{\partial X_A} \right) = 0.$$

The last term in eq. (58) can be cast in a tractable form by applying the orthonormality condition:

$$\begin{aligned} \sum_{\mu\nu} C_{\mu i} S_{\mu\nu} C_{\nu j} &= \delta_{ij} \\ &= \sum_{k,l=\text{fr}} \sum_{\alpha\beta} \sigma_i U_{ik} C_{\alpha k}^0 R_{\mu\alpha}^k S_{\mu\nu} R_{\nu\beta}^l C_{\beta l}^0 U_{lj} \sigma_j \end{aligned} \quad (59)$$

This implies the following:

$$\sum_{\alpha\beta} \sum_{\mu\nu} C_{\alpha k}^0 R_{\mu\alpha}^k S_{\mu\nu} R_{\nu\beta}^l C_{\beta l}^0 = \sum_{i=\text{fr}} U_{ki} \sigma_i^{-2} U_{il} \quad (60)$$

Taking derivatives for this last equation gives:

$$\begin{aligned} & \frac{\partial(\sum_i U_{ki} \sigma_i^{-2} U_{il})}{\partial X_A} \\ &= \sum_i \left(\frac{\partial U_{ki}}{\partial X_A} U_{il} + U_{ki} \frac{\partial U_{il}}{\partial X_A} \right) \\ & \quad + \sum_i U_{ki} \frac{\partial(\sigma_i^{-2})}{\partial X_A} U_{il} \end{aligned} \quad (61)$$

$$\begin{aligned} &= \sum_{\alpha} \sum_{\mu\nu} C_{\alpha k}^0 \frac{\partial R_{\mu\alpha}^k}{\partial X_A} S_{\mu\nu} C_{\nu l}^0 \\ & \quad + \sum_{\beta} \sum_{\mu\nu} C_{\mu k}^0 S_{\mu\nu} \frac{\partial R_{\nu\beta}^k}{\partial X_A} C_{\beta l}^0 + \sum_{\mu\nu} C_{\mu k}^0 S_{\mu\nu} C_{\nu l}^0 \end{aligned} \quad (62)$$

Now, because $\sum_i ((\partial U_{ki}/\partial X_A) U_{il} + U_{ki} (\partial U_{il}/\partial X_A)) = 0$ and $(\partial(\sigma_i^{-2})/\partial X_A) = -2(\partial\sigma_i/\partial X_A) = -(\partial(\sigma_i^2)/\partial X_A)$:

$$\begin{aligned} & \sum_i U_{ki} \frac{\partial(\sigma_i^2)}{\partial X_A} U_{il} \\ &= - \sum_i U_{ki} \frac{\partial(\sigma_i^{-2})}{\partial X_A} U_{il} \\ &= - \sum_{\alpha} \sum_{\mu\nu} C_{\alpha k}^0 \frac{\partial R_{\mu\alpha}^k}{\partial X_A} S_{\mu\nu} C_{\nu l}^0 \\ & \quad - \sum_{\beta} \sum_{\mu\nu} C_{\mu k}^0 S_{\mu\nu} \frac{\partial R_{\nu\beta}^k}{\partial X_A} C_{\beta l}^0 - \sum_{\mu\nu} C_{\mu k}^0 S_{\mu\nu} C_{\nu l}^0 \end{aligned} \quad (63)$$

Substituting this expression into eq. (57), changing some subscript labels, and rearranging some terms results in the following:

$$\begin{aligned} \frac{\partial P_{\mu\nu}}{\partial X_A} &= \sum_{i=\text{fr}} \sum_{\alpha} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i}^0 C_{\nu i}^0 + \sum_{i=\text{fr}} \sum_{\alpha} C_{\mu i}^0 \frac{\partial R_{\nu\alpha}^i}{\partial X_A} C_{\alpha i}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} C_{\mu i}^0 \frac{\partial R_{\gamma\alpha}^i}{\partial X_A} C_{\alpha i}^0 S_{\gamma\delta} C_{\delta j}^0 C_{\nu j}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 S_{\gamma\delta} \frac{\partial R_{\delta\alpha}^i}{\partial X_A} C_{\alpha j}^0 C_{\nu j}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j}^0 C_{\nu j}^0 \end{aligned} \quad (64)$$

Now that an expression for $(\partial P_{\mu\nu}/\partial X_A)$ has been obtained, it can be rewritten as:

$$\begin{aligned} \frac{\partial P_{\mu\nu}}{\partial X_A} &= \sum_{a=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} C_{\nu a}^0 + \sum_{a=\text{fr}} C_{\mu a}^0 \frac{\partial C_{\nu a}}{\partial X_A} \\ &= \sum_{i=\text{fr}} \sum_{\alpha} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i}^0 C_{\nu i}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} C_{\mu i}^0 \frac{\partial R_{\gamma\alpha}^i}{\partial X_A} C_{\alpha i}^0 S_{\gamma\delta} C_{\delta j}^0 C_{\nu j}^0 \\ & \quad - \frac{1}{2} \sum_{i,j=\text{fr}} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j}^0 C_{\nu j}^0 \\ & \quad + \sum_{i=\text{fr}} \sum_{\alpha} C_{\mu i}^0 \frac{\partial R_{\nu\alpha}^i}{\partial X_A} C_{\alpha i}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 S_{\gamma\delta} \frac{\partial R_{\delta\alpha}^i}{\partial X_A} C_{\alpha j}^0 C_{\nu j}^0 \\ & \quad - \frac{1}{2} \sum_{i,j=\text{fr}} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j}^0 C_{\nu j}^0 \end{aligned} \quad (65)$$

and thus:

$$\begin{aligned} \frac{\partial C_{\mu a}}{\partial X_A} C_{\nu a}^0 &= \sum_{i=\text{fr}} \sum_{\alpha} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i}^0 C_{\nu i}^0 \\ & \quad - \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} C_{\mu i}^0 \frac{\partial R_{\gamma\alpha}^i}{\partial X_A} C_{\alpha i}^0 S_{\gamma\delta} C_{\delta j}^0 C_{\nu j}^0 \\ & \quad - \frac{1}{2} \sum_{i,j=\text{fr}} \sum_{\gamma\delta} C_{\mu i}^0 C_{\gamma i}^0 \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j}^0 C_{\nu j}^0 \end{aligned} \quad (66)$$

This can now be used in the first term on the right-hand side of eq. (53) to get:

$$\begin{aligned} & 4 \sum_{\mu\nu} \sum_{a=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} F_{\mu\nu}^* C_{\nu a} \\ &= 4 \sum_{i=\text{fr}} \sum_{\mu\nu} \sum_{\alpha} \frac{\partial R_{\mu\alpha}^i}{\partial X_A} C_{\alpha i}^0 F_{\mu\nu}^* C_{\nu i} \\ & \quad - 4 \sum_{i,j=\text{fr}} \sum_{\alpha} \sum_{\gamma\delta} \epsilon_{ij}^* \frac{\partial R_{\gamma\alpha}^i}{\partial X_A} C_{\alpha i}^0 S_{\gamma\delta} C_{\delta j}^0 \\ & \quad - 2 \sum_{i,j=\text{fr}} \sum_{\gamma\delta} \epsilon_{ij}^* C_{\gamma i}^0 \frac{\partial S_{\gamma\delta}}{\partial X_A} C_{\delta j}^0 \end{aligned} \quad (67)$$

where ϵ_{ij}^* is defined as $\sum_{\mu\nu} C_{\mu i}^0 F_{\mu\nu}^* C_{\nu j}^0$. Also, the superscripts have been dropped from C^0 because it is understood that this is the coefficient matrix at the current geometry.

The next task is to transform the second term on the right-hand side of eq. (53) into a viable form. The first step is to use the modified Roothaan equations [eqs. (7) and (8)] to obtain:

$$\begin{aligned}
 & 4 \sum_{\mu\nu} \sum_{a=\text{non}} \frac{\partial C_{\mu a}}{\partial X_A} F_{\mu\nu}^* C_{\nu a} \\
 &= 4 \sum_{\mu\nu} \sum_{i=\text{non}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu i} \epsilon_i^* \\
 &+ 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu j} \epsilon_{ij}^* \\
 &+ 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{virt.}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu j} \epsilon_{ij}^* \quad (68)
 \end{aligned}$$

The first term on the right-hand side can be substituted through use of:

$$\begin{aligned}
 & \sum_{\mu\nu} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu b} + \sum_{\mu\nu} C_{\mu a} \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu b} \\
 &+ \sum_{\mu\nu} C_{\mu a} S_{\mu\nu} \frac{\partial C_{\nu b}}{\partial X_A} = 0 \quad (69)
 \end{aligned}$$

while the second term can be rewritten using:

$$\begin{aligned}
 & 2 \sum_{\mu\nu} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu a} = - \sum_{\mu\nu} C_{\mu a} C_{\nu a} \frac{\partial S_{\mu\nu}}{\partial X_A} \\
 & \quad (\text{when } a = b) \quad (70)
 \end{aligned}$$

The following is thus obtained:

$$\begin{aligned}
 & 4 \sum_{\mu\nu} \sum_{i=\text{non}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu i} \epsilon_i^* \\
 &= -2 \sum_{\mu\nu} \sum_{i=\text{non}} C_{\mu i} \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu i} \epsilon_i^* \quad (71) \\
 & 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu j} \epsilon_{ij}^* \\
 &= -4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} C_{\mu i} \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu j} \epsilon_{ij}^* \\
 &- 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} C_{\mu i} S_{\mu\nu} \frac{\partial C_{\nu j}}{\partial X_A} \epsilon_{ij}^* \quad (72)
 \end{aligned}$$

The derivative $(\partial C_{\nu j} / \partial X_A)$ in the last term of eq. (72) can be calculated explicitly using eq. (54):

$$\begin{aligned}
 \frac{\partial C_{\nu j}}{\partial X_A} &= \sum_{\alpha} \frac{\partial R_{\nu\alpha}^j}{\partial X_A} C_{\alpha j} + \sum_{k=\text{fr}} C_{\nu k} \frac{\partial U_{kj}}{\partial X_A} \\
 &+ \sum_{k=\text{fr}} C_{\nu k} U_{kj} \frac{\partial \sigma_j}{\partial X_A} \quad (73)
 \end{aligned}$$

and, thus, substitution yields:

$$\begin{aligned}
 & 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} \frac{\partial C_{\mu a}}{\partial X_A} S_{\mu\nu} C_{\nu j} \epsilon_{ij}^* \\
 &= -4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} C_{\mu i} \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu j} \epsilon_{ij}^* \\
 &- 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j=\text{fr}} \sum_{\alpha} C_{\mu i} S_{\mu\nu} \frac{\partial R_{\nu\alpha}^j}{\partial X_A} C_{\alpha j} \epsilon_{ij}^* \\
 &- 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j,k=\text{fr}} C_{\mu i} S_{\mu\nu} C_{\nu k} \frac{\partial U_{kj}}{\partial X_A} \epsilon_{ij}^* \\
 &- 4 \sum_{\mu\nu} \sum_{i=\text{non}} \sum_{j,k=\text{fr}} C_{\mu i} S_{\mu\nu} C_{\nu k} U_{kj} \frac{\partial \sigma_j}{\partial X_A} \epsilon_{ij}^* \quad (74)
 \end{aligned}$$

The last two terms, however, are both equal to zero due to the orthogonality of the nonfrozen orbitals to the frozen orbitals ($\sum_{\mu\nu} C_{\mu i} S_{\mu\nu} C_{\nu k} = 0$ for $i = \text{nonfrozen}$ and $k = \text{frozen}$).

The last term from eq. (68) remains to be recast in a tractable form. This transformation begins with the observation that the frozen virtuals will be orthogonal to the nonfrozen occupied orbitals in the new geometry both before and after the next SCF procedure. Therefore, the following relations among the derivatives can be formulated:

$$\begin{aligned}
 & \sum_{\mu\nu} \frac{\partial C_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 + \sum_{\mu\nu} C_{\mu i}^0 \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu j}^0 \\
 &+ \sum_{\mu\nu} C_{\mu i}^0 S_{\mu\nu} \frac{\partial C_{\nu j}}{\partial X_A} = 0 \quad (75)
 \end{aligned}$$

$$\begin{aligned}
 & \sum_{\mu\nu} \frac{\partial \bar{C}_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 + \sum_{\mu\nu} C_{\mu i}^0 \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu j}^0 \\
 &+ \sum_{\mu\nu} C_{\mu i}^0 S_{\mu\nu} \frac{\partial C_{\nu j}}{\partial X_A} = 0 \quad (76)
 \end{aligned}$$

where \bar{C} indicates the initial nonfrozen occupied coefficients in the new geometry before performing SCF calculations (i.e., right after re-orthonormalization) and the superscripts on C^0 again

emphasize the current geometry (constant) coefficients. Note that there is no distinction between the frozen virtual orbitals before the next SCF procedure and the frozen virtuals after SCF as they will remain frozen once changed to reflect the new geometry. Thus, eqs. (75) and (76) imply the following:

$$\sum_{\mu\nu} \frac{\partial C_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 = \sum_{\mu\nu} \frac{\partial \bar{C}_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 \quad (77)$$

By noting that the nonfrozen occupied after re-orthonormalization will be linear combinations of the current frozen orbitals projected onto the basis functions allowed for use by the nonfrozen orbitals and the current geometry nonfrozen orbitals, the following can be written:

$$\begin{aligned} \sum_{\mu\nu} \frac{\partial \bar{C}_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 &= \sum_{\mu\nu} \sum_{f=\text{fr}} A_{if} \bar{C}_{\mu f}^0 S_{\mu\nu} C_{\nu j}^0 \\ &+ \sum_{\mu\nu} \sum_{n=\text{non}} A_{in} C_{\mu n}^0 S_{\mu\nu} C_{\nu j}^0 \end{aligned} \quad (78)$$

where $\bar{C}_{\mu f}^0$ indicates a coefficient for the frozen orbital f projected onto the basis functions allowed for nonfrozen orbitals ($\bar{C}_{\mu f}^0 = C_{\mu f}^0$ for $\mu = \text{allowed}$ and $\bar{C}_{\mu f}^0 = 0$ for $\mu = \text{nonallowed}$), and \mathbf{A} is the matrix of coefficients for the linear combinations determined during orthonormalization. Due to the orthogonality of the nonfrozen orbitals to the frozen virtuals, the last term in the previous equation is zero ($\sum_{\mu\nu} C_{\mu n}^0 S_{\mu\nu} C_{\nu j}^0 = 0$), whereas the orthogonality of the nonfrozen orbitals to the frozen occupied allows for the following:

$$\begin{aligned} 0 &= \sum_{\mu\nu} \frac{\partial \bar{C}_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu f}^0 + \sum_{\mu\nu} C_{\mu i}^0 \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu f}^0 \\ &+ \sum_{\mu\nu} C_{\mu i}^0 S_{\mu\nu} \frac{\partial C_{\nu f}}{\partial X_A} \quad (79) \\ &= \sum_{\mu\nu} \sum_{g=\text{fr}} A_{ig} \bar{C}_{\mu g}^0 S_{\mu\nu} C_{\nu f}^0 + \sum_{\mu\nu} C_{\mu i}^0 \frac{\partial S_{\mu\nu}}{\partial X_A} C_{\nu f}^0 \\ &+ \sum_{\mu\nu} \sum_{\alpha} C_{\mu i}^0 S_{\mu\nu} \frac{\partial R_{\nu\alpha}^f}{\partial X_A} C_{\alpha f}^0 \quad (80) \end{aligned}$$

Substitutions from eqs. (68) and (78) were used in eq. (80), keeping in mind that anytime a factor analogous to $\sum_{\mu\nu} C_{\mu i}^0 S_{\mu\nu} C_{\nu j}^0$ occurs with $i = \text{frozen}$ and $j = \text{nonfrozen}$, that term will be equal to zero.

If the following definitions are made:

$$\tilde{\mathcal{S}}^\dagger \equiv \tilde{\mathbf{C}}^{0\dagger} \mathbf{S} \mathbf{C}^0 \quad (81)$$

$$\tilde{\zeta}_i^{\dagger\dagger} \equiv \mathbf{C}^{0\dagger} \mathbf{S}' \mathbf{C}^0, \left(S'_{\mu\nu} = \frac{\partial S_{\mu\nu}}{\partial X_A} \right) \quad (82)$$

$$\tilde{\zeta}_i^{R\dagger} \equiv \mathbf{C}^{0\dagger} \mathbf{S} \mathbf{R}'^f \mathbf{C}^0, \left(R'_{\nu\alpha}^f = \frac{\partial R_{\nu\alpha}^f}{\partial X_A} \right) \quad (83)$$

then eq. (80) yields the following set of linear equations to be solved for $\tilde{a}_i^{\dagger\dagger}$ ($\tilde{a}_i^{\dagger\dagger}$ is a row of \mathbf{A}):

$$\tilde{a}_i^{\dagger\dagger} \tilde{\mathcal{S}}^\dagger = -\tilde{\zeta}_i^{\dagger\dagger} - \tilde{\zeta}_i^{R\dagger} \quad (84)$$

This system of linear equations can be solved by using the singular value decomposition of $\tilde{\mathcal{S}}^\dagger = \mathbf{V} \mathbf{W} \mathbf{U}^\dagger$ ($\tilde{\mathcal{S}}^{\dagger-1} = \mathbf{U} \mathbf{W}^{-1} \mathbf{V}^\dagger$):

$$\tilde{a}_i^{\dagger\dagger} = -\tilde{\zeta}_i^{\dagger\dagger} \mathbf{U} \mathbf{W}^{-1} \mathbf{V}^\dagger - \tilde{\zeta}_i^{R\dagger} \mathbf{U} \mathbf{W}^{-1} \mathbf{V}^\dagger \quad (85)$$

where \mathbf{U} and \mathbf{V} are orthogonal matrices, and \mathbf{W} is a diagonal matrix.

Now, using the resulting expression for $\tilde{a}_i^{\dagger\dagger}$ in the last equation and the definitions in eqs. (81)–(83), eq. (78) can be rewritten as:

$$\begin{aligned} \sum_{\mu\nu} \frac{\partial \bar{C}_{\mu i}}{\partial X_A} S_{\mu\nu} C_{\nu j}^0 &= - \sum_{f,g,h=\text{fr}} \sum_{\mu\nu} \sum_{\alpha\beta} C_{\alpha i} \\ &\times \frac{\partial S_{\alpha\beta}}{\partial X_A} C_{\beta g} U_{gh} w_h^{-1} V_{fh} \bar{C}_{\mu f} S_{\mu\nu} C_{\nu j} \\ &- \sum_{f,g,h=\text{fr}} \sum_{\mu\nu} \sum_{\alpha\beta\gamma} C_{\alpha i} S_{\alpha\beta} \\ &\times \frac{\partial R_{\beta\gamma}^g}{\partial X_A} C_{\gamma g} U_{gh} w_h^{-1} V_{fh} \bar{C}_{\mu f} S_{\mu\nu} C_{\nu j} \end{aligned} \quad (86)$$

where the 0 superscripts have again been dropped.

Finally, a complete expression can be depicted for the QM/MM gradients. First, the following definition is made so that the result is more compact:

$$\omega_{gi} \equiv \sum_{j=\text{fr. virt.}} \sum_{f,h=\text{fr}} \sum_{\mu\nu} U_{gh} w_h^{-1} V_{fh} \bar{C}_{\mu f} S_{\mu\nu} C_{\nu j} \epsilon_{ji}^* \quad (87)$$

Now, by using this definition, along with eqs. (53), (67), (68), (74), and (86) in eq. (22), eq. (24) is produced.

DERIVATION OF ROTATION MATRIX DERIVATIVES

As for finding the rotation matrix itself, the first step in finding the derivatives for the rotation

matrix elements is to find the derivatives for the 3×3 matrix, $\mathbf{P}_{\text{tot}} = \mathbf{P}_2 \mathbf{P}_1$. \mathbf{P}_1 here is the matrix that aligns the reference frame in the current geometry (defined as before in Fig. 6 by \mathbf{d} , \mathbf{c} , and \mathbf{a}) in the \hat{x} , \hat{y} , and \hat{z} directions, respectively. Because this matrix depends only on the current geometry, it remains constant. \mathbf{P}_2 , however, is not constant as \mathbf{P}_2^{-1} is what transforms the vectors \mathbf{d}' , \mathbf{c}' , and \mathbf{a}' , which define the reference frame of the new geometry, to be aligned in the \hat{x} , \hat{y} , and \hat{z} directions, respectively. Thus, $(\partial/\partial X_A)(\mathbf{P}_{\text{tot}}) = (\partial/\partial X_A)(\mathbf{P}_2)\mathbf{P}_1$, where:

$$\mathbf{P}_1 = \begin{pmatrix} \frac{d_x}{d} & \frac{d_y}{d} & \frac{d_z}{d} \\ \frac{c_x}{c} & \frac{c_y}{c} & \frac{c_z}{c} \\ \frac{a_x}{a} & \frac{a_y}{a} & \frac{a_z}{a} \end{pmatrix}; \quad (88)$$

$$\frac{\partial}{\partial X_A}(\mathbf{P}_2) = \frac{\partial}{\partial X_A} \begin{pmatrix} \frac{d'_x}{d'} & \frac{c'_x}{c'} & \frac{a'_x}{a'} \\ \frac{d'_y}{d'} & \frac{c'_y}{c'} & \frac{a'_y}{a'} \\ \frac{d'_z}{d'} & \frac{c'_z}{c'} & \frac{a'_z}{a'} \end{pmatrix}$$

To find $(\partial/\partial X_A)(\mathbf{P}_2)$, it is first noted that \mathbf{P}_2 only depends on the coordinates for atoms 1, 2, and 3 (\mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3), as defined in Figure 7. Thus, for $X_A \neq x_1$, y_1 , z_1 , x_2 , y_2 , z_2 , x_3 , y_3 , or z_3 , $(\partial/\partial X_A)(\mathbf{P}_2) = 0$. It must be stated here that, to make formulations easier for the derivatives, the expressions will include the $\mathbf{P}_{\text{tor}}(\phi_i)$ matrices implicitly in the \mathbf{P}_{tot} matrix by setting atom 3 to atom i for the other atoms i connected to the bond adjacent atoms 1 and 2 when X_A is a coordinate for atom i (i.e., when $i = A$). Note that it may be

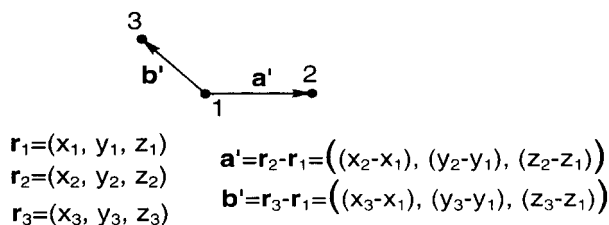


FIGURE 7. Atoms used to find rotational matrix derivatives are defined here as atoms 1, 2, and 3, with their respective positional vectors \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 .

necessary to switch atoms 1 and 2 to keep the new atom 3 on the same side of the frozen bond (axis \mathbf{a}) as atom 1. Thus, \mathbf{P}_{tot} will depend on atom A .

Now, looking at eq. (88), it is observed that we need the derivatives $(\partial/\partial X_A)(\mathbf{a}'/d')$, $(\partial/\partial X_A)(\mathbf{c}'/c')$, and $(\partial/\partial X_A)(\mathbf{d}'/d')$. Using the chain rule, these derivatives can be expressed as:

$$\frac{\partial}{\partial X_A} \left(\frac{\mathbf{a}'}{d'} \right) = \frac{1}{d'} \frac{\partial \mathbf{a}'}{\partial X_A} - \frac{\mathbf{a}'}{d'^2} \frac{\partial d'}{\partial X_A} \quad (89)$$

$$\frac{\partial}{\partial X_A} \left(\frac{\mathbf{c}'}{c'} \right) = \frac{1}{c'} \frac{\partial \mathbf{c}'}{\partial X_A} - \frac{\mathbf{c}'}{c'^2} \frac{\partial c'}{\partial X_A} \quad (90)$$

$$\frac{\partial}{\partial X_A} \left(\frac{\mathbf{d}'}{d'} \right) = \frac{1}{d'} \frac{\partial \mathbf{d}'}{\partial X_A} - \frac{\mathbf{d}'}{d'^2} \frac{\partial d'}{\partial X_A} \quad (91)$$

Keeping in mind that $\mathbf{c}' = \mathbf{a}' \times \mathbf{b}'$ and $\mathbf{d}' = \mathbf{c}' \times \mathbf{a}'$, the following can be written using the chain rule:

$$\frac{\partial \mathbf{a}'}{\partial X_A} = \frac{\mathbf{a}'}{d'} \cdot \frac{\partial \mathbf{a}'}{\partial X_A} \quad (92)$$

$$\frac{\partial \mathbf{c}'}{\partial X_A} = \frac{\partial \mathbf{a}'}{\partial X_A} \times \mathbf{b}' + \mathbf{a}' \times \frac{\partial \mathbf{b}'}{\partial X_A} \quad (93)$$

$$\frac{\partial \mathbf{c}'}{\partial X_A} = \frac{\mathbf{c}'}{c'} \cdot \frac{\partial \mathbf{c}'}{\partial X_A} \quad (94)$$

$$\frac{\partial \mathbf{d}'}{\partial X_A} = \frac{\partial \mathbf{c}'}{\partial X_A} \times \mathbf{a}' + \mathbf{c}' \times \frac{\partial \mathbf{a}'}{\partial X_A} \quad (95)$$

$$\frac{\partial \mathbf{d}'}{\partial X_A} = \frac{\mathbf{d}'}{d'} \cdot \frac{\partial \mathbf{d}'}{\partial X_A} \quad (96)$$

All the above derivatives can be calculated once the derivatives $(\partial \mathbf{a}'/\partial X_A)$ and $(\partial \mathbf{b}'/\partial X_A)$ are determined. Referring to the expressions for \mathbf{a}' and \mathbf{b}' (shown in Fig. 7), the derivatives can be computed easily and are displayed in the following table:

X_A	$\frac{\partial \mathbf{a}'}{\partial X_A}$	$\frac{\partial \mathbf{b}'}{\partial X_A}$
x_1	$(-1, 0, 0)$	$(-1, 0, 0)$
y_1	$(0, -1, 0)$	$(0, -1, 0)$
z_1	$(0, 0, -1)$	$(0, 0, -1)$
x_2	$(1, 0, 0)$	$(0, 0, 0)$
y_2	$(0, 1, 0)$	$(0, 0, 0)$
z_2	$(0, 0, 1)$	$(0, 0, 0)$
x_3	$(0, 0, 0)$	$(1, 0, 0)$
y_3	$(0, 0, 0)$	$(0, 1, 0)$
z_3	$(0, 0, 0)$	$(0, 0, 1)$

Now all the necessary derivatives can be evaluated, with the derivatives for $\mathbf{P}_{\text{tot}}^A$ given by:

$$\frac{\partial \mathbf{P}_{\text{tot}}^A}{\partial X_A} = \begin{pmatrix} \frac{1}{d'} \frac{\partial d'_x}{\partial X_A} - \frac{d'_x}{d'^2} \frac{\partial d'}{\partial X_A} & \frac{1}{c'} \frac{\partial c'_x}{\partial X_A} - \frac{c'_x}{c'^2} \frac{\partial c'}{\partial X_A} & \frac{1}{a'} \frac{\partial a'_x}{\partial X_A} - \frac{a'_x}{a'^2} \frac{\partial a'}{\partial X_A} \\ \frac{1}{d'} \frac{\partial d'_y}{\partial X_A} - \frac{d'_y}{d'^2} \frac{\partial d'}{\partial X_A} & \frac{1}{c'} \frac{\partial c'_y}{\partial X_A} - \frac{c'_y}{c'^2} \frac{\partial c'}{\partial X_A} & \frac{1}{a'} \frac{\partial a'_y}{\partial X_A} - \frac{a'_y}{a'^2} \frac{\partial a'}{\partial X_A} \\ \frac{1}{d'} \frac{\partial d'_z}{\partial X_A} - \frac{d'_z}{d'^2} \frac{\partial d'}{\partial X_A} & \frac{1}{c'} \frac{\partial c'_z}{\partial X_A} - \frac{c'_z}{c'^2} \frac{\partial c'}{\partial X_A} & \frac{1}{a'} \frac{\partial a'_z}{\partial X_A} - \frac{a'_z}{a'^2} \frac{\partial a'}{\partial X_A} \end{pmatrix} \mathbf{P}_1^A \quad (97)$$

where the label A emphasizes the dependence on atom A .

As was done earlier, the rest of the blocks of the rotational matrix derivatives can be determined from $(\partial \mathbf{P}_{\text{tot}}^A / \partial X_A)$. Because s functions are spherically symmetric and invariant to rotations:

$$\frac{\partial \mathbf{S}}{\partial X_A} = (\mathbf{S}'_{1,1}) = (0) \quad (98)$$

The derivative matrix for p functions is just the matrix of derivatives for the rotation of atom A

from the current geometry to a new geometry:

$$\frac{\partial \mathbf{P}^A}{\partial X_A} = \begin{pmatrix} P'_{1,1} & P'_{1,2} & P'_{1,3} \\ P'_{2,1} & P'_{2,2} & P'_{2,3} \\ P'_{3,1} & P'_{3,2} & P'_{3,3} \end{pmatrix} = \frac{\partial \mathbf{P}_{\text{tot}}^A}{\partial X_A} \quad (99)$$

The matrix of rotational derivatives for d functions can be derived by differentiating the \mathbf{D}' matrix found earlier [eq. (34)] using the chain rule and keeping in mind that \mathbf{P}^0 (the \mathbf{P}_{tot} rotational matrix evaluated at the current geometry itself) is simply the identity matrix $\mathbf{1}$:

$$\frac{\partial \mathbf{D}^A}{\partial X_A} = \begin{pmatrix} 2P'_{1,1} & 0 & 0 & 2\sqrt{3}P'_{1,2} & 2\sqrt{3}P'_{1,3} & 0 \\ 0 & 2P'_{2,2} & 0 & 2\sqrt{3}P'_{2,1} & 0 & 2\sqrt{3}P'_{2,3} \\ 0 & 0 & 0 & 2P'_{3,3} & 2\sqrt{3}P'_{3,1} & 2\sqrt{3}P'_{3,2} \\ \frac{2}{\sqrt{3}}P'_{2,1} & \frac{2}{\sqrt{3}}P'_{1,2} & 0 & P'_{1,1} + P'_{2,2} & P'_{2,3} & P'_{1,3} \\ \frac{2}{\sqrt{3}}P'_{3,1} & 0 & \frac{2}{\sqrt{3}}P'_{1,3} & P'_{2,3} & P'_{1,1} + P'_{3,3} & P'_{1,2} \\ 0 & \frac{2}{\sqrt{3}}P'_{3,2} & \frac{2}{\sqrt{3}}P'_{2,3} & P'_{3,1} & P'_{2,1} & P'_{2,2} + P'_{3,3} \end{pmatrix} \quad (100)$$

Derivatives of rotational matrices for higher angular momentum blocks will not be presented here. Once the rotational matrix derivatives have been computed for all individual blocks of basis functions on all relevant atoms, the matrix $(\partial \mathbf{R}^k / \partial X_A)$ (where k is a label for which the frozen orbital is being rotated) is constructed from these matrices in the same way that \mathbf{R} was constructed earlier.

References

1. Warshel, A.; Levitt, M. *J Mol Biol* 1976, 103, 227.
2. Singh, U. C.; Kollman, P. A. *J Comput Chem* 1986, 7, 718.
3. Bash, P. A.; Field, M. J.; Karplus, M. *J Am Chem Soc* 1987, 109, 8092.
4. Gao, J.; Xia, X. *Science* 1992, 258, 631.
5. Luzhkov, V.; Warshel, A. *J Comput Chem* 1992, 13, 199.
6. Stanton, R. V.; Hartsough, D. S.; Merz, K. M. *J Phys Chem* 1993, 97, 11868.
7. Tuñón, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F.; Rivail, J.-L. *J Comput Chem* 1996, 17, 19.
8. Bakowies, D.; Thiel, W. *J Phys Chem* 1996, 100, 10580.
9. Maseras, F.; Morokuma, K. *J Comput Chem* 1995, 16, 1170.
10. Field, M. J.; Bash, P. A.; Karplus, M. *J Comput Chem* 1990, 11, 700.
11. Eurenium, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int J Quantum Chem* 1996, 60, 1189.

12. Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* 1991, 30, 5826.
13. Karplus, M.; Evanseck, J. D.; Joseph, D.; Bash, P. A.; Field, M. *Faraday Discuss* 1992, 93, 239.
14. Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J Chem Soc Perkin Trans 2* 1991, 225, 1819.
15. Hartsough, D. S.; Merz, K. M. *J Phys Chem* 1995, 99, 384.
16. Chatfield, D. C.; Brooks, B. R. *J Am Chem Soc* 1995, 117, 5561.
17. Mulholland, A. J.; Richards, W. G. *Protein* 1997, 27, 9.
18. Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J Comput Chem* 1994, 15, 269.
19. Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J Phys Chem* 1998, 102, 4714.
20. Assfeld, X.; Rivail, J.-L. *Chem Phys Lett* 1996, 263, 100.
21. Humbel, S.; Sieber, S.; Morokuma, K. *J Chem Phys* 1996, 105, 1959.
22. Matsubara, T.; Sieber, S.; Morokuma, K. *Int J Quantum Chem* 1996, 60, 1101.
23. Foster, J. M.; Boys, S. F. *Rev Mod Phys* 1960, 32, 300.
24. Schrödinger, Inc., Portland, OR, JAGUAR v3.5, 1998.
25. The IMPACT Group, IMPACT, v7.0, 1996.
26. Jorgensen, W. L.; Tirado-Rives, J. *J Am Chem Soc* 1988, 110, 1657.
27. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J Am Chem Soc* 1996, 118, 11225.
28. Schlegel, H. B. *Theor Chim Acta* 1984, 66, 333.
29. Fletcher, R. *Practical Methods of Optimization*, Vol. 1; Wiley: New York, 1980.
30. Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J Am Chem Soc* 1996, 119, 5908.
31. Pulay, P.; Sæbø, S.; Meyer, W. *J Chem Phys* 1984, 81, 1901.
32. Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J Chem Phys* 1995, 103, 1481.
33. Dunning, T. H., Jr. *J Chem Phys* 1989, 90, 1007.
34. Hohenberg, P.; Kohn, W. *Phys Rev* 1964, 136, B864.
35. Kohn, W.; Sham, L. J. *Phys Rev* 1965, 140, A1133.
36. Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*, No. 16 (International Series of Monographs on Chemistry); Oxford University Press: Oxford, 1989.
37. Andzelm, J. DGauss: Density functional–Gaussian approach. Implementation and Applications. In: Labanowski, J. K.; Andzelm, J. W., eds. *Density Functional Methods in Chemistry*; Springer: New York, 1991.